

JOHNSON GRANT

IN-61

7307

P-58

# **EXPERT SYSTEM VERIFICATION AND VALIDATION STUDY**

## **Phase 2: Requirements Identification**

### **Delivery 1 - Updated Survey Report**

N91-20790

Unclass  
0007307

G3/61

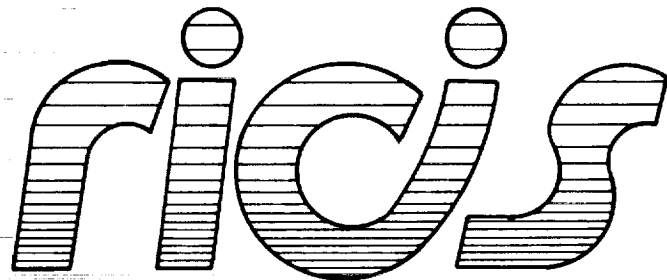
(NASA-CR-168109) EXPERT SYSTEM VERIFICATION  
AND VALIDATION STUDY, PHASE 2: REQUIREMENTS  
IDENTIFICATION, DELIVERY 1: UPDATED SURVEY  
REPORT (Houston Univ.) 58 p CSCL 09B

**International Business Machines Corporation**

**December 14, 1990**

**Cooperative Agreement NCC 9-16  
Research Activity No. AI.16**

**NASA Johnson Space Center  
Information Systems Directorate  
Information Technology Division**



**Research Institute for Computing and Information Systems  
University of Houston - Clear Lake**

**T · E · C · H · N · I · C · A · L      R · E · P · O · R · T**

## *The RICIS Concept*

The University of Houston-Clear Lake established the Research Institute for Computing and Information systems in 1986 to encourage NASA Johnson Space Center and local industry to actively support research in the computing and information sciences. As part of this endeavor, UH-Clear Lake proposed a partnership with JSC to jointly define and manage an integrated program of research in advanced data processing technology needed for JSC's main missions, including administrative, engineering and science responsibilities. JSC agreed and entered into a three-year cooperative agreement with UH-Clear Lake beginning in May, 1986, to jointly plan and execute such research through RICIS. Additionally, under Cooperative Agreement NCC 9-16, computing and educational facilities are shared by the two institutions to conduct the research.

The mission of RICIS is to conduct, coordinate and disseminate research on computing and information systems among researchers, sponsors and users from UH-Clear Lake, NASA/JSC, and other research organizations. Within UH-Clear Lake, the mission is being implemented through interdisciplinary involvement of faculty and students from each of the four schools: Business, Education, Human Sciences and Humanities, and Natural and Applied Sciences.

Other research organizations are involved via the "gateway" concept. UH-Clear Lake establishes relationships with other universities and research organizations, having common research interests, to provide additional sources of expertise to conduct needed research.

A major role of RICIS is to find the best match of sponsors, researchers and research objectives to advance knowledge in the computing and information sciences. Working jointly with NASA/JSC, RICIS advises on research needs, recommends principals for conducting the research, provides technical and administrative support to coordinate the research, and integrates technical results into the cooperative goals of UH-Clear Lake and NASA/JSC.

***EXPERT SYSTEM VERIFICATION AND  
VALIDATION STUDY***

***Phase 2: Requirements Identification***

***Delivery 1 - Updated Survey Report***

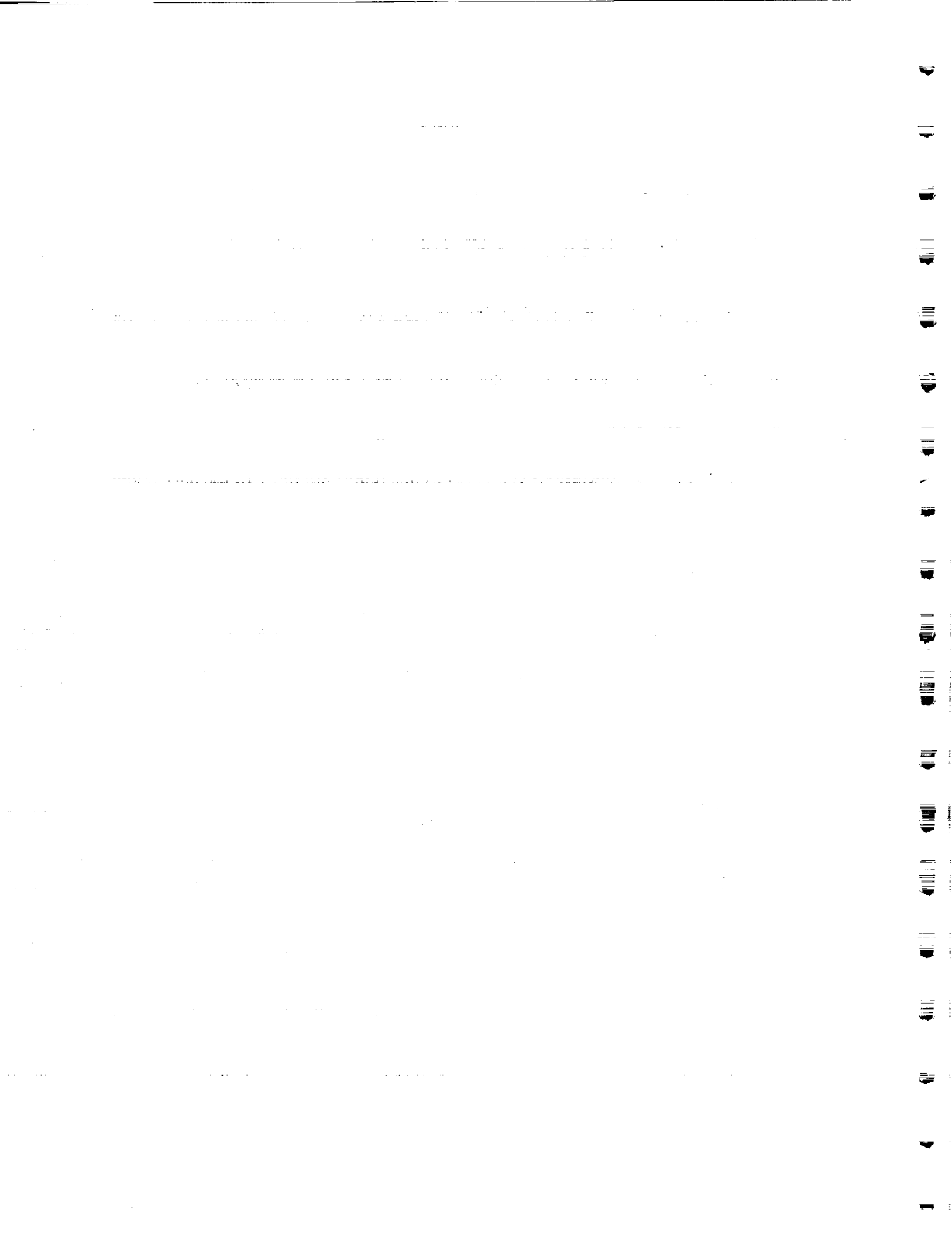


## Preface

This research was conducted under the auspices of the Research Institute for Computing and Information Systems by The International Business Machines Corporation. Dr. Terry Feagin and Dr. Ted Leibfried served as RICIS research representatives for this research activity.

Funding has been provided by Information Technology Division, Information Systems Directorate, NASA/JSC through Cooperative Agreement NCC 9-16 between NASA Johnson Space Center and the University of Houston-Clear Lake. The NASA technical monitor for this activity was Chris Culbert, of the Software Technology Branch, Information Technology Division, Information Technology Directorate, NASA/JSC.

The views and conclusions contained in this report are those of the author and should not be interpreted as representative of the official policies, either express or implied, of NASA or the United States Government.



**Expert System Verification and Validation Study  
RICIS Contract #069**

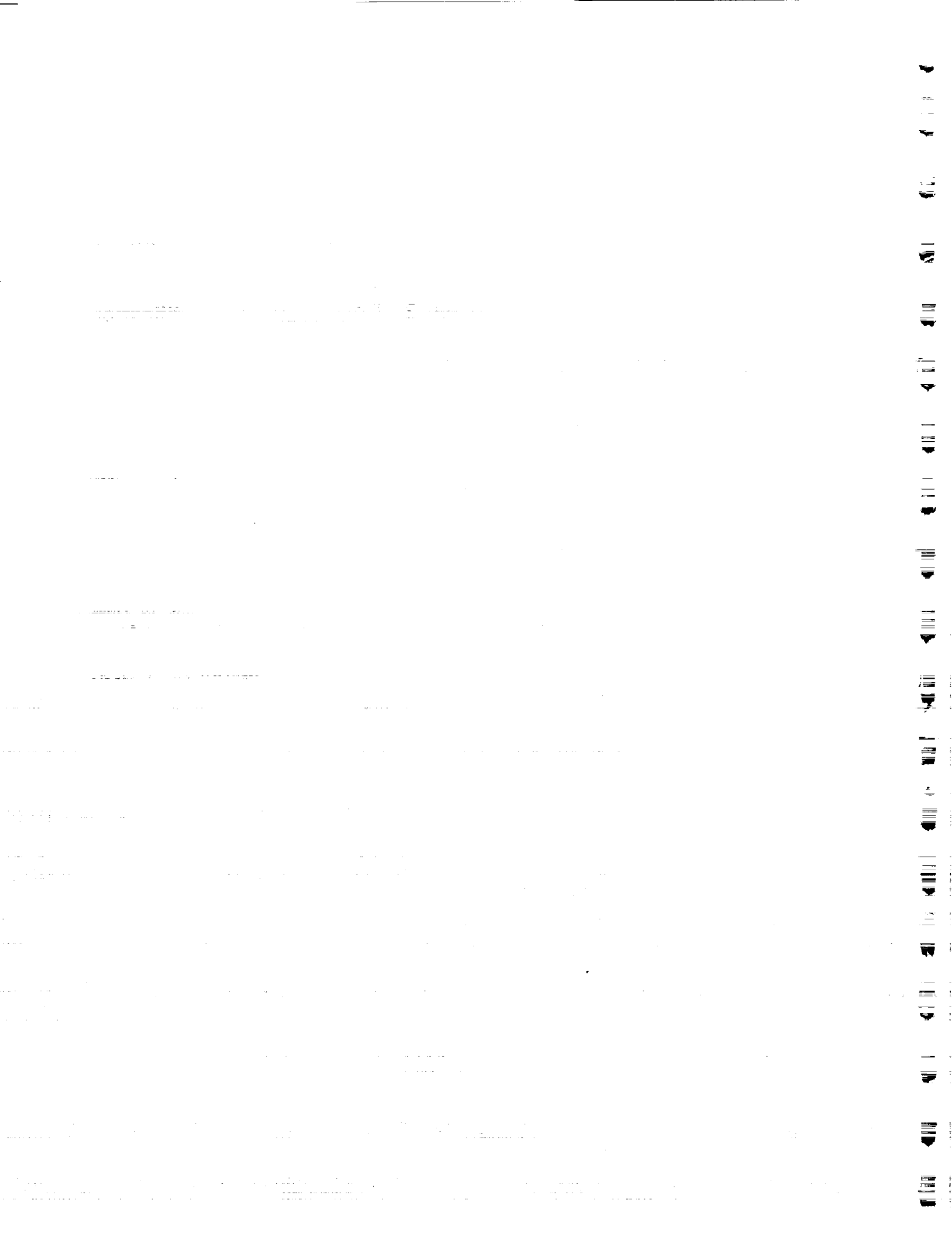
**Phase 2 - Requirements Identification  
Delivery 1 - Updated Survey Report**

December 14, 1990

**IBM**

3700 Bay Area Blvd.  
Houston, Tx. 77058

**Updated Survey Report**

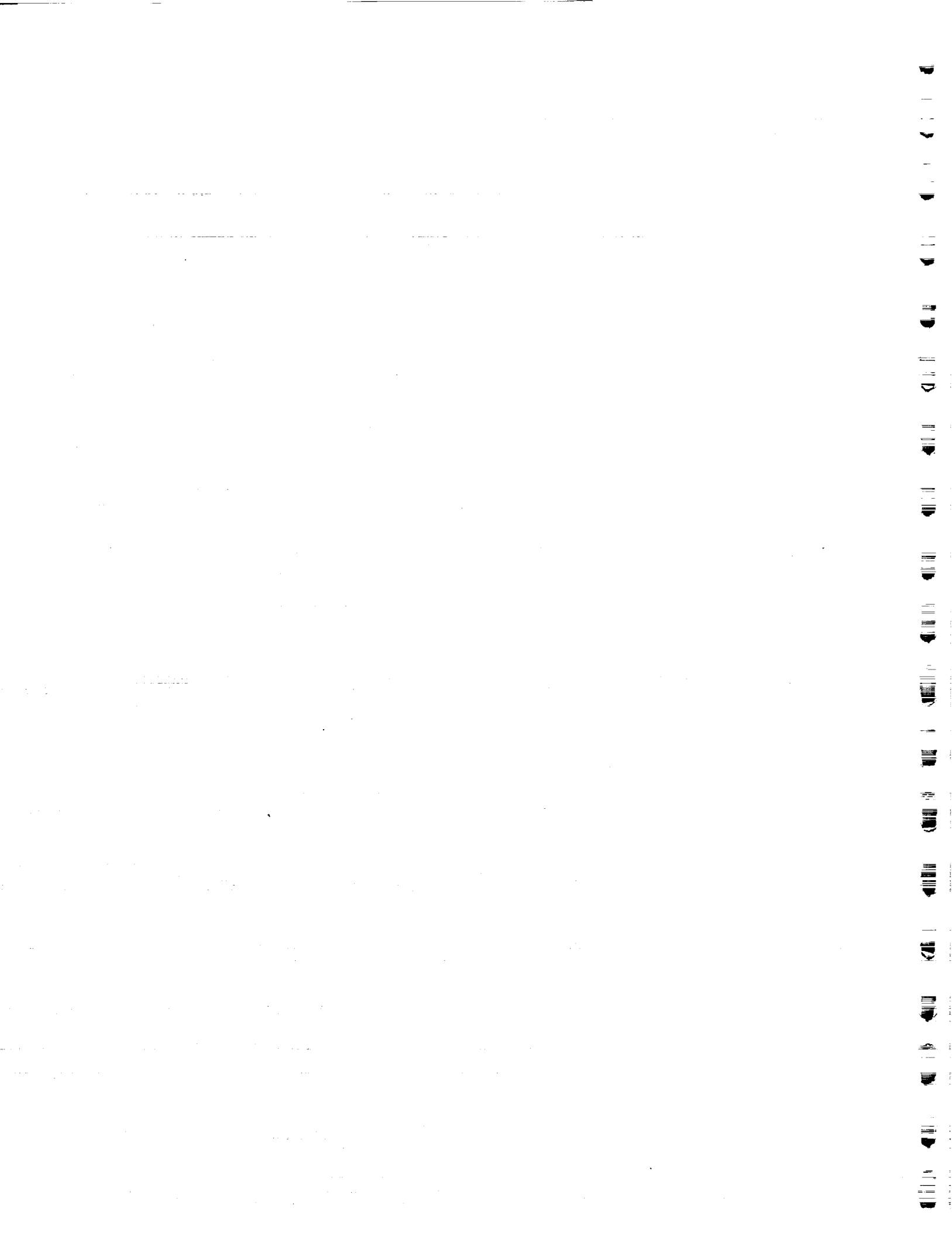




---

## Preface

This document constitutes the first delivery, "Updated Survey Report," of the four deliveries scheduled for the second phase of RICIS contract 069, "Verification and Validation of Expert Systems Study." This deliverable is an update to the "Revised Final Report," delivered on October 31, 1990, which was the final delivery of the first phase of this contract.



## **Contents**

<b>Background</b>	<b>1</b>
<b>Executive Summary</b>	<b>2</b>
<b>Survey Rationale</b>	<b>3</b>
Purpose of the Questionnaires	4
Purpose of the Interviews	4
<b>Survey Administration</b>	<b>5</b>
<b>Survey Questionnaires</b>	<b>6</b>
Information Gathered	6
Human Factors	7
<b>Summary of Results</b>	<b>8</b>
General information	9
Performance Criteria	10
Requirements Definition	12
Development Information	13
V&V Activities Performed	15
V&V Issues Encountered	17
<b>Summary of Interview Results</b>	<b>20</b>
<b>Recommendations</b>	<b>23</b>
Direct Recommendations	23
Inferred Recommendations	26
<b>Appendix A. Detailed results</b>	<b>29</b>
Summary of Developers Responses (part 1)	29
Summary of Developers Responses (part 2)	33
Summary of Users Responses	37
<b>Appendix B. Expert Systems Evaluation Questionnaire (Developer)</b>	<b>38</b>
<b>Appendix C. Expert Systems Evaluation Questionnaire (User)</b>	<b>46</b>



## **Background**

The purpose of this deliverable is to report the state-of-the-practice in Verification and Validation (V&V) of Expert Systems (ESs) on current NASA and Industry applications. This is the first task of a series which has the ultimate purpose of ensuring that adequate ES V&V tools and techniques are available for Space Station Knowledge Based Systems development.

The strategy for determining the state-of-the-practice is to check how well each of the known ES V&V issues are being addressed and to what extent they have impacted the development of Expert Systems.

**Note:** This task does not attempt to prove or disprove whether Verification and Validation can or should be performed on Expert Systems. It is accepted that Verification and Validation should be applied to all software systems, including Expert Systems.

---

## Executive Summary

Data from over sixty Expert System (ES) projects was collected through a written survey and/or interviews. Forty basic questions were asked, ranging over a variety of general topics such as the size of the ES and the difficulty in specifying requirements. However, all the questions were designed to gather information about different aspects of V&V. Significant results include the following points (see "Summary of Results" on page 8 for the actual percentages):

1. In most cases, the ES was expected to be at least as accurate as the expert but often the ES was less accurate.
2. All users estimated the ES to be less accurate than expected while half the developers estimated the ES to be less accurate than expected.
3. Less than half the systems had a requirements document.
4. On average a quarter of the developers time was spent on V&V.
5. While developers thought evaluating an expert system was of average difficulty, users unanimously thought it was hard.
6. All V&V techniques were used, with each technique being relied upon, by at least one project, as the sole V&V technique used.
7. The most often cited V&V problems were test coverage determination, knowledge validation, and problem complexity.

Based on an analysis of the survey results, several recommendations were formulated. These recommendations are:

1. Develop suggested V&V requirements for ESs, that is, standard and guidelines V&V of ESs at each stage of development.
2. Address the test coverage determination, knowledge validation, and problem complexity issues.
3. Develop ways to make knowledge bases more easily modularized and easier to understand.
4. Address the configuration management of expert systems.
5. Develop criteria to classify an ES by intended use so that V&V requirements can be tailored to different types of ESs.
6. Investigate ways to assist an expert in analyzing a knowledge base, possibly either through the use of analysis tools or higher level representations.

## Survey Rationale

It is widely claimed that Expert Systems have been not been subject to the same level of Verification and Validation as traditionally developed software. Some people feel that this lack of V&V continues because of a "vicious circle," where nobody requires expert system V&V, so nobody does it. Consequently, since nobody knows how to do it, nobody requires it. There are two major reasons why the V&V process has not been documented: lack of a single life-cycle model, and technical differences between traditional software and expert systems.

Most expert system development life-cycles rely on iterative prototypes to develop the system behavior. This approach does not lead to methodical capture and documentation of the expected system behavior. Documented expectations, traditionally captured in a requirements document, are essential in the V&V process: you can't do testing if you don't know what to test for! One goal of this survey is to understand how the expected behavior of current expert systems is communicated and evaluated, even if a formal requirements document was not developed.

Expert Systems are typically composed of three parts: the knowledge base (KB), the inference engine, and the interface code between the inference engine and the peripheral devices (terminals, sensors, effectors, users, etc.). The inference engine and interface code are simply traditional software and should currently be V&Ved by accepted practices. This survey will help determine if these parts are V&Ved or whether, since they are part of an expert system, V&V is overlooked.

The knowledge base is the only part of the Expert System that raises new and unique issues. A set of the *possible* issues are:

### Issues primarily due to use of nonprocedural languages

- Understandability and readability to support inspections
- Testing coverage
- Standard validation tests for inference engines
- Real-time performance analysis

### Issues due to heuristic knowledge (difficulty in organizing)

- Knowledge validation
- Modularity Design

### Issues primarily due to solving new complex problems

- Requirements
- Certification

### Other issues

- Uncertainty Analysis
- Inheritance Process Test and Analysis
- Configuration Management

One of the purposes of this survey is to find out if these identified *possible* issues actually cause problems in practice, and if so, how the issues are being handled.

---

## Purpose of the Questionnaires

Some of the information for this survey can be captured fairly easily and is accomplished through use of a questionnaire. The information captured this way includes:

- Application information - What kind of problem does the system address?, What are the performance goals?
- Expertise information - What was the relationship between the developers and expert(s)?, What is the performance level of the expert?
- Development information - How was the system developed?, How big is the system?
- Evaluation information - How was the system evaluated?
- Performance information - How important is good performance?, How well is the ES performing?

---

## Purpose of the Interviews

The questionnaire answers lead to an additional set of questions involving the V&V issues described earlier. The additional questions are greatly affected by the answers provided in top questionnaire, so it would be more efficient to derive the information through direct interviews than to generate a large number of secondary questionnaires. The interviews attempt to uncover:

- the real issues involved in ES V&V (in comparison with the known *possible* issues outlined above).
- what is being done currently to address V&V (inspections, path testing, testing by the expert).
- what makes users trust the ESs, if the ESs are indeed trusted.
- what problems, unique to ESs, were encountered and possibly addressed during development and test.

The interviews are also required because we expect that some people will not fill out the questionnaires.



## Survey Administration

This survey was designed so that the majority of the information would be gained from direct interviews with people involved in ES projects. Several people from each project, including developers, users, and managers, were interviewed to get a realistic view of the projects.

Several other activities were undertaken, both before and after the interview activity, to ensure that the results of the survey reflected the actual "state-of-the-practice". These activities included:

### Identifying candidate ES projects

A list of projects to be contacted was created. The list included projects at NASA and IBM as well as projects from fields outside of the space industry.

### Developing survey questionnaire(s)

To improve the chances of getting meaningful data from the questionnaire activity, separate questionnaires were developed for developers and users. Each questionnaire includes a question to indicate if the answers are from a manager or non-manager. Questionnaires are listed in Appendix B, "Expert Systems Evaluation Questionnaire (Developer)" on page 38 and Appendix C, "Expert Systems Evaluation Questionnaire (User)" on page 46.

### Evaluating returned questionnaires

Each questionnaire was evaluated to determine if project interviews would uncover more information. If a project was to be interviewed, the questionnaire results provided guidance on which topics would be the most useful to explore.

### Summarizing interview/questionnaire results

The summarized results of the questionnaire/interview activities are presented in section "Summary of Results" on page 8.

### Recommendations

Recommendations for further action, based on the information in "Summary of Results" on page 8 are provided in section "Recommendations" on page 23.

ORIGINAL PAGE IS  
OF POOR QUALITY

## Survey Questionnaires

Different versions of the questionnaire were developed for developers and users of the expert system. In addition, responses were expected to be different between managers and non-managers, so an indication is included on each questionnaire.

### Information Gathered

Several types of information are captured by the questionnaire. Each question in the questionnaire addresses at least one of the previous types of information. For each type of information, the subtopics and questions which provide information are listed. The question numbers are noted as (development question, user question). Questions not available on a questionnaire are indicated by a "-".

#### General Information

Describes the general properties of the expert system, including the name (1, 1), a short description (4, 4), field of the problem (5, 5), and the type of problem to be solved (6, 6). Also captured are whether the survey taker was a manager (2, 2).

#### Performance Criteria

A major expertise issue is performance (probability that the results given are correct); specifically performance of the experts (10, 9), expected performance of the system (11, 10), and actual performance of the system (12, 11). Related to the performance issue is the amount of the problem space that the ES is expected to cover (8, 7), and that it actually covers (9, 8).

#### Requirements Definition

Requirements definition information includes how the requirements are documented (13, -), the difficulty in determining the requirements (14, -), and the availability of the expert(s) to resolve requirements issues during development (17, -). Influencing the performance issue is the number of experts (15, -), and whether the experts agree on the results obtained from the system (16, 21). It may also be useful to know if the expert (-, 12) and/or the developer(s) (18, 13) are part of the user organization.

#### Development Information

Development information that we are concerned with includes the development life-cycle used (19, -), and what languages and tools were used to develop the system (20, -). The size of the system (22, -), the total effort required for development, (29, -), and the effort required to develop the different parts of the ES (21, -) indicate the difficulty of the development effort. The sensitivity of the system (24, -) will influence the difficulty of future maintenance activities.

#### V&V Activities Performed

The major information to be captured during this task is the current state-of-the-practice for V&V of ESs, including the kinds of V&V being attempted, both during (28, -) and after (33, 20) development, and how much of the development effort was spent on V&V (30, -). Detailed information is also gathered for V&V activities for Knowledge Structures (25, -), the Inference Engine (26, -), and the Interface Code (27, -).

Information about the difficulty of the V&V effort (35, 22), whether a separate group performed V&V, (31, -) and how much effort was expended on the independent V&V (32, 19), is also gathered.

Whether the system is operational or prototype (3, 3), and the criticality of the system (37, 15) have an affect on the amount of V&V activities performed.

#### **V&V Issues Encountered**

If the state-of-the-practice is to be improved, the major issues that need to be addressed must be identified. One question (36, 23) directly asks whether each the known issues was actually encountered. Additional questions find out more information about specific issues, including the existence of certainty factors (7, -), whether configuration management was performed (34, -), and the difficulty of implementing the expertise through the Knowledge Structures (23, -). User acceptance is the ultimate test of the V&V activities. The comparison between expected system use (39, 17) and actual system use (40, 18), the perceived reliability of the system (38, 16), and why the user is convinced that the system produces correct results (-, 14) are all indicators of user acceptance.

---

## **Human Factors**

The questionnaires were designed to capture as much accurate information as possible. In an effort to accomplish this, the following human factors issues were taken into account:

#### **Questions should be understandable**

Questions should have as few "technical" terms as possible to avoid confusion due to local usage. For questions that must have technical content, be sure to provide sufficient explanation.

#### **Choices worded positively**

Negatively worded choices may not get selected because the responder may feel there is something wrong with it.

#### **Meaningful questions**

The responder should feel that there is some purpose to the question.

#### **Make use of fill-in-the-blank questions**

The responder should not have to fill in long responses. Some questions can not have all possible responses enumerated, so the user should be able to specify his own choice.

ORIGINAL PAGE IS  
OF POOR QUALITY

## Summary of Results

The survey results are summarized in the following sections. The results are organized according to the type of information, as organized in "Information Gathered" on page 6. The percentages in parentheses correspond to the results from the developer and user questionnaire, respectively. If the question is not in one of the questionnaires, the position is filled with a '-'.

### General Information

Most of the respondents were involved with Expert Systems which perform Diagnosis (45%,80%), primarily in the Aerospace field (46%,100%). The survey respondents were predominantly involved with development (93%).

### Performance Criteria

(37%,40%) estimated an actual accuracy of less than 90% and (48%,60%) estimated an accuracy of less than 95%. Most (60%,40%) estimated the problem space coverage between 60% and 95%. In comparing the accuracy of the expert and the expert system, most expected the expert system to at least as accurate as the expert (78%,80%) while the expert system often was estimated to be less accurate than expected (49%,100%) and less accurate than the expert (44%,80%). Note that the results show that users more often (than developers) cited the system as being less accurate than expert and less accurate than expected.

### Requirements Definition

(75%,-) indicated that expert consultation was a basis for determining the behavior of the system. More revealing is that (52%,-) said there were not any documented requirements and (43%,-) indicated that prototypes or similar tools were used for requirements.

(40%,-) had medium difficulty in generating requirements while (35%,-) said they were hard and (25%,-) said they were easy. (58%,-) of developers had a high level of contact with experts during development.

### Development Information

The most frequent (40%,-) Life-Cycle model used is the Cyclic Model (repetition of Requirements, Design, Rule Generation, and Prototyping until done); however, (22%,-) of the respondents stated that no model was followed. Most development was done with an Expert System shell (CLIPS and others), and the predominant Interface Code was C and LISP. Applications were reasonably large, requiring an average of 33 person/months to develop. Developed systems were not reported to be particularly sensitive to change; (77%,-) said changes only occasionally caused an unexpected behavior.

### V&V Activities Performed

Most V&V activities relied on comparison with expected results and expert checking. Typically, (24%,-) of the development effort was spent on V&V. While developers seemed to feel V&V was of medium difficulty, users unanimously agreed that it was hard; (34%,0%) said it was medium while (27%,100%) said it was hard and (33%,0%) said it was easy; (5%,0%) said it was impossible. Of significant interest is the fact that each V&V technique was used as the sole V&V technique in at least one project. Also, in general, there was wide ranging uses of V&V tech-

niques. (39%,20%) of the respondents indicated that the ES was a prototype system.

#### V&V Issues Encountered

The known issues most often cited as problems were: test coverage determination (50%,75%), knowledge validation (44%,75%), problem complexity (39%,40%), and real-time performance analysis (40%,25%). (Note that as a whole, the developers ranking of the issues agreed with the users ranking of the issues). The least cited problem was analysis of certainty factors (only seven respondents indicated that certainty factors were used). Every known issue was cited by at least one respondent.

Configuration management practices are reported to be an issue for many participants, regardless of whether the system was operational or a prototype.

The expected system use varied widely (3-2000), while actual system use was relatively good (less than half of the respondents provided information, suggesting that actual use was much lower than reported).

The following sections list the results from each individual question. The total number of responses is given for each question along with the number of times each choice was selected ( given to the left of the choice).

---

## General information

The questions for the name of the ES, and the short description are not reported.

### Field of the Problem

Question Numbers: 5, 5

Total Responses: 70

What field does the problem belong to?

- 35 Aerospace
- \_4 Financial
- \_2 Information Systems
- \_8 Hardware
- \_6 Manufacturing
- \_2 Marketing
- \_ Medical
- \_1 Personnel
- \_2 Research
- \_1 Service
- \_4 Software
- \_5 Other

### Type of Problem Solved

Question Numbers: 6, 6

Total Responses: 70

Which of the following items best describes the kind of problem the Expert System addresses? Please indicate primary purpose with a '\*' and check all other applicable purposes (if any).

**Note:** The number of times the choice was selected as primary purpose is given in parentheses after the number of times the choice was selected.

- 13 (11) Design - Configuring objects under constraints
- 11 (0) Repair - Executing plans to administer prescribed remedies
- 11 (5) Control - Governing overall system behavior
- 16 (5) Planning - Designing actions
- 34 (23) Diagnosis - Inferring system malfunctions from observables
- 11 (1) Debugging - Prescribing remedies for malfunctions
- 16 (3) Prediction - Inferring likely consequences of given situations
- 23 (8) Monitoring - Comparing observations to expected outcomes
- 12 (1) Instruction - Diagnosing, debugging, and repairing behavior
- 15 (5) Interpretation - Inferring situation descriptions from sensor data
- 5 (2) Classification - Categorizing objects by properties
- 3 ( ) Others

## Role on Project

Question Numbers: 2, 2

Total Responses: 70

Were you a developer of the Expert System the manager of the, development organization, a user of the Expert System, or the manager of a department which uses the Expert System?

- 42 Developer of Expert System
- 6 Manager of Expert System development organization
- 17 Other Development
- 4 User of the Expert System
- Manager of a department using the Expert System
- 1 Other User

---

## Performance Criteria

### Performance of the Experts

Question Numbers: 10, 9

Total Responses: 70

If human experts currently perform (or previously performed) the task, how often is the expert(s) expected to give the correct answer?

- 2 Task not performed by human
- 17 "Correct" defined by expert
- 19 > 99%
- 16 95% to 99%
- 4 90% to 95%
- 4 80% to 90%
- 1 60% to 80%
- 40% to 60%
- 4 Other (2 - 100%)
- 3 I don't know

### Expected Performance of the System

Question Numbers: 11, 10

Total Responses: 70

How often is the Expert System expected to provide the correct answer?

- 22 100%
- 16 > 99%
- 9 95% to 99%
- 10 90% to 95%
- 4 80% to 90%
- 3 60% to 80%
- 40% to 60%
- 1 Other
- 5 I don't know

### Actual Performance of the System

Question Numbers: 12, 11

Total Responses: 68

What is *your estimate* of how often the Expert System *actually* provides the correct answer?

- 11 100%
- 11 > 99%
- 12 95% to 99%
- 10 90% to 95%
- 8 80% to 90%
- 5 60% to 80%
- 1 40% to 60%
- 3 Other (< 40%)
- 7 I don't know

### Expected Problem Space Coverage

Question Numbers: 8, 7

Total Responses: 70

How much of the problem space is the Expert System expected to cover?

- 15 100%
- 12 > 99%
- 6 95% to 99%
- 7 90% to 95%
- 13 80% to 90%
- 4 60% to 80%
- 4 40% to 60%
- 4 Other
- 5 I don't know

### Actual Problem Space Coverage

Question Numbers: 9, 8

Total Responses: 70

What is *your estimate* of the problem space coverage *actually* provided by the Expert System?

- 4 100%

- ☐ 3 > 99%
- ☐ 8 95% to 99%
- ☐ 3 90% to 95%
- ☐ 14 80% to 90%
- ☐ 19 60% to 80%
- ☐ 8 40% to 60%
- ☐ 7 Other (1 - 5%)
- ☐ 8 I don't know

---

## Requirements Definition

### Requirements Format

Question Numbers: 13, -  
Total Responses: 62

What was the basis for determining how the system was to behave? Please indicate the primary basis with a '\*' and check all other applicable basis (if any).

**Note:** The number of times the choice was selected as primary basis is given in parentheses after the number of times the choice was selected.

- ☐ 12 (☐ 4) A pre-existing document
- ☐ 19 (☐ 4) A requirements document completed as part of development.
- ☐ 6 (☐ ) Some other developed document
- ☐ 27 (☐ 4) A prototype of the system
- ☐ 49 (38) Expert consultation
- ☐ 6 (☐ )

### Requirements Difficulty

Question Numbers: 14, -  
Total Responses: 63

How difficult was it to develop the original concept of what the system was supposed to do?

- ☐ 7 Trivial
- ☐ 15 Easy
- ☐ 25 Medium
- ☐ 15 Hard
- ☐ 1 Impossible

### Availability of the Expert(s)

Question Numbers: 17, -  
Total Responses: 53

If the system was not developed by the expert, how much interaction was there between the expert(s) and the development team?

- ☐ 6 System was developed by expert
- ☐ 10 Constant
- ☐ 15 Frequent
- ☐ 17 Regular
- ☐ 5 Occasional
- ☐ None



## Number of Experts

Question Numbers: 15, -  
Total Responses: 64

Was more than one expert consulted during the development of the system?

10 System was developed by expert  
\_6 Single expert  
30 Multiple experts with lead  
12 Committee of experts  
\_6 Other

## Agreement Among Experts

Question Numbers: 16, 21  
Total Responses: 61

If more than one expert was available for consulting, how often did the experts agree on what results the Expert System was supposed to provide?

\_6 A single expert was involved  
11 Always agree  
44 Agree 75% of the time (range 30%-99%)

## Expert in User Organization

Question Numbers: -, 12  
Total Responses: 5

Was the expert(s) a member of the user organization?

\_5 Yes  
\_ No  
\_ User organization provided some expertise

## Developers in User Organization

Question Numbers: 18, 13  
Total Responses: 69

Was the developer(s) of the Expert System part of the user organization?

25 Yes  
31 No  
13 Some development provided by user organization

---

## Development Information

### Development Life-Cycle Used

Question Numbers: 19, -  
Total Responses: 58

Please indicate which development model was used for developing the Expert System.

\_5 Requirements gathering preceded Design, Implementation, and Test (Traditional waterfall life-cycle).  
12 Requirements gathered before development of a prototype. A second requirements activity preceded Design, Implementation, and Test.

25 Repetition of the Requirements, Design, Rule Generation, and Prototyping phases until production system (final prototype) was developed.  
14 No effort was made to follow a particular model.  
\_2 Other

## Languages and Tools Used

Question Numbers: 20, -  
Total Responses: 64

What was the primary language/tool for the knowledge structures ?

**Note:** The most frequent languages/tools are reported after the choice as: "frequency - language/tool."

Knowledge Structures (17 - ESE, 13 - CLIPS, 10 - LISP, others)

## Size of the System

Question Numbers: 22, -  
Total Responses: 39

Since Knowledge Bases can be written using several type of Knowledge Structures, please indicate how many of the following structures were used. If another type of structure was used, please describe it and how many were used.

**Note:** The number of times that a value was given for each choice is provided in parentheses followed by the average value for that response. The range of the responses is given in parentheses after each choice.

(35) 235 Rules (range 30-1000)  
(15) 872 Frames (range 1-10000)  
(10) 248 Facts (range 50-800)  
(15) 121 Parameters (range 20-400)  
( 2) 8K Statements (2K - 16K)

## Total Development Effort

Question Numbers: 29, -  
Total Responses: 57

How much effort was expended in developing the system, including evaluation activities performed by the developers? 33 (range 1-200) person/months.

## Detailed Development Effort

Question Numbers: 21, -  
Total Responses: 64

What percentage of the total development effort was dedicated to each part of the Expert System?

61 % Knowledge Structures  
8 % Inference Engine  
31 % Interface Code

## System Sensitivity

Question Numbers: 24, -

Total Responses: 64

When changes were made to the knowledge structures, how often did some unexpected result occur?

- \_5 Never
- 44 Occasionally
- \_9 Frequently
- \_5 Usually
- \_1 Always

---

## V&V Activities Performed

### V&V Activities during development

Question Numbers: 28, -

Total Responses: 63

What testing activities were performed on the executing system? (indicate any that apply)

- \_2 No evaluation was performed
- 38 Checked by expert(s)
- 32 Compared with expected results
- 28 Structural testing (e.g. cover all rules)
- 18 Other

### V&V Activities after development

Question Numbers: 33, 20

Total Responses: 47

What testing activities were performed on the executing system before the system was delivered to the users? (indicate any that apply)

- \_1 No evaluation was performed
- 33 Checked by expert(s)
- 39 Compared with expected results
- 29 User acceptance
- 16 System run in parallel
- \_5 Other

### Development effort was spent on V&V

Question Numbers: 30, -

Total Responses: 62

How much of the development effort was spent on evaluation? 24 % (range 2%-80%)

## V&V of Knowledge Structures

Question Numbers: 25, -

Total Responses: 65

What evaluation activities were performed on the Knowledge Structures? (indicate any that apply)

- ☐ 3 No evaluation was performed
- ☐ 28 Desk checking
- ☐ 15 Formal inspections
- ☐ 42 Checked by expert(s)
- ☐ 39 Structural testing (e.g. cover all rules)
- ☐ 9 Other

## V&V of Inference Engine

Question Numbers: 26, -

Total Responses: 35

What evaluation activities were performed on the Inference Engine? (indicate any that apply)

- ☐ 17 No evaluation was performed (ES shell was used)
- ☐ 2 No evaluation was performed
- ☐ 3 Desk checking
- ☐ 10 Formal inspections
- ☐ 5 Structural testing
- ☐ Other

## V&V of Interface Code

Question Numbers: 27, -

Total Responses: 58

What evaluation activities were performed on the Interface Code? (indicate any that apply)

- ☐ 7 No evaluation was performed
- ☐ 25 Desk checking
- ☐ 12 Formal inspections
- ☐ 29 Structural testing (branch or path)
- ☐ 18 Experts
- ☐ Other

## Difficulty of V&V

Question Numbers: 35, 22

Total Responses: 67

Compared to conventional software testing efforts, how difficult was the evaluation of the Expert System?

- ☐ 3 Trivial
- ☐ 16 Easy
- ☐ 20 Medium
- ☐ 20 Hard
- ☐ 3 Impossible
- ☐ 4 No evaluation was done

## Separate V&V group

Question Numbers: 31, -

Total Responses: 62

Did a separate organization evaluate the Expert System before it was delivered to the users?

15 Yes, there was a separate evaluation organization.

47 No, there was not a separate evaluation organization.

## Independent V&V Effort

Question Numbers: 32, 19

Total Responses: 11

If there was a separate evaluation team, how much effort was expended by the team in evaluating the correctness of the Expert System?

(11) 3 (range 1-7) person/months reported by developers

(3) 16 (range 3-24) person/months reported by users

## Operational or Prototype System

Question Numbers: 3, 3

Total Responses: 70

Is the Expert System operational or is it a prototype?

42 Operational system

25 Prototype system

\_3 Operational prototype (write in)

## System Criticality

Question Numbers: 37, 15

Total Responses: 69

How reliable is the Expert System required to be?

\_7 Trusted with human life

15 Trusted with mission objectives

31 As reliable as the expert

17 Assists the expert

19 Assists the user

\_ Other

---

## V&V Issues Encountered

### Known Issues Actually Encountered

Question Numbers: 36, 23

Total Responses: 66

Many people feel that some development issues are more of a problem with Expert Systems than with conventional systems. Which (if any) of the following were problems during implementation or test of this Expert System?

13 Understandability and readability of knowledge structures

34 Determining test coverage for knowledge structures

19 Modularity/Design of knowledge structures

- 30 Knowledge validation
- \_6 Analysis of Certainty Factors
- \_8 Validating the inference engine
- 26 Real-time performance analysis
- 26 Complexity of the Problem
- 14 Certification
- \_9 Configuration Management
- \_6 Other

## Certainty Factors

Question Numbers: 7, -  
Total Responses: 64

Does the Expert System include certainty factors?

- \_7 Yes
- 54 No
- \_3 I don't know

## Configuration Management

Question Numbers: 34, -  
Total Responses: 45

How were changes to the Expert System distributed to the users?

- \_5 User updated system at developer's direction
- 18 Developers made changes to users' system
- \_1 Untested system distributed to users
- 22 Tested system distributed to the users
- \_3 Configuration management group distributes system
- \_1 Other

## Expertise Implementation Difficulty

Question Numbers: 23, -  
Total Responses: 62

Aside from any difficulties in developing the original concept, how difficult was it to express the behavior (through the Knowledge Structures) of the expert?

- \_3 Trivial
- 16 Easy
- 20 Medium
- 20 Hard
- \_3 Impossible

## Expected System Use

Question Numbers: 39, 17  
Total Responses: 50

How many people are expected to make use of the Expert System? 219 (range 1-2000)

## Perceived System Reliability

Question Numbers: 38, 16

Total Responses: 68

Does the Expert System seem to be more reliable or less reliable than conventional systems that are in use?

- ☐ 9 Significantly more reliable
- ☐ 16 More reliable
- ☐ 3 Slightly more reliable
- ☐ 19 Similar reliability
- ☐ 2 Slightly less reliable
- ☐ 1 Less reliable
- ☐ Significantly less reliable
- ☐ 14 No comparison is available
- ☐ 4 I don't know

## User Trust

Question Numbers: -, 14

Total Responses: 5

Why do you believe the results that the system gives?

- ☐ 1 Expert says it is correct
- ☐ 3 Participated in evaluation
- ☐ Someone I trust did evaluation
- ☐ 5 Personal use and checking
- ☐ 1 User acceptance
- ☐ I don't trust the results
- ☐ Other

---

## Summary of Interview Results

In addition to acquiring written responses to the survey questions, interviews were performed to gather additional data and to clarify questions concerning the written responses. Additional information from these interviews are summarized in this section.

**Structural Testing:** Based on the survey results, a commonly used evaluation approach was the use of structural testing. This was surprising because it was felt that structural testing was relatively difficult to apply to expert systems. From the interviews, we learned that although some projects did attempt to measure the actual test coverage (i.e., percentage of rules executed during testing) many others did not actually measure the coverage. Instead, they attempted to develop test cases that would cover all of the knowledge base (or at least the important parts) but made no attempt to measure how well the knowledge base was actually covered. Also, there appeared to be no attempt to cover interactions between knowledge base elements (e.g., rule interactions); each element was tested as if it were an independent piece of the knowledge base. Some knowledge base developers felt that more formal structural testing would be too much effort and would hinder the development process too much. In conclusion, it seemed that, although structural testing was used, it was a very weak form of structural testing (at least compared to, say, branch coverage in procedural software testing).

**Experts Developing Expert Systems:** It appeared that the expert was heavily relied upon to aid in evaluation of the knowledge base; this subject was probed more deeply during the interviews. It seems that a close interaction between the expert and the knowledge base developer was mandatory to successfully develop an expert system. This is not a surprising result and it has been discussed at length in the literature. However, it was surprising to learn that many knowledge base developers feel that this interaction is so important that they think the best approach is simply to have the expert develop the system. However, one non-programmer interviewee, who felt that his group was being successful at having experts develop their own systems, also thought that this approach would have to altered to some extent in order to be successful at the more sophisticated types of expert systems that they would be developing in the future.

**Requirements Writing and the Conventional Software Life-Cycle:** It was anticipated that expert systems were being developed using a much more iterative and less structured life-cycle than the conventional and rigid waterfall model. And, although the subject of life-cycle models was not intentionally addressed during the interviews, it often came up when discussing requirements. It seems that several respondents associated "requirements" with the conventional waterfall model and they felt very strongly that the conventional approaches to software development, such as the waterfall model, were much too formal and structured for expert systems development - that is, it would be disastrous to apply them to expert systems. Though for some, this feeling extended to requirements, others simply used a different approach to requirements. For example, in some cases, requirements were not written because it was felt that a requirements document was a formally written paper document that needed to be "approved" before development could proceed. While in other cases, an iterative prototyping development effort took place and was followed by documenting system requirements; these requirements were then used to test the system to ensure that it worked as everyone thought it (supposedly) did.



**Prototypes vs. Operational Systems:** Although we attempted to get respondents to state that their system was either "a prototype" or "operational," we received indications that this distinction was not easy to make, in practice. For example, responses included "it is both a prototype and operational," or "it is an operational prototype," or "it is just a prototype but we have many users." It seems that some systems are originally intended to be a prototype but become used operationally. Some intentionally approach the development of an operational system by first developing a "prototype" and once the prototype is "certified," it is considered "operational." However, there is a danger that a prototype will be used as if it were operational. Some have made efforts to ensure that a system that was only intended to be a prototype system was not accidentally relied upon in an operational setting.

**Real-Time Performance Analysis:** It was intended that "real-time performance analysis" would refer to the ability to predict the response time for an expert system. That is, the ability to analyze the time performance of the system. However, from the interviews we learned that many interpreted "real-time performance analysis" to mean the ability to get the system to run as fast as desired/necessary.

## Issues Independent of A System Being an Expert System

An important, but difficult, aspect of analyzing expert system development methodology is distinguishing properties of expert systems that are significantly different from properties of conventional software. This is also an important aspect of the analysis of this survey of V&V issues. Several comments appeared to be due more to factors other than the fact that the system being developed was an "expert" system. The interviews helped clarify this issue which the remainder of this section discusses.

**Extensive Use of Prototyping and Rapid Development:** The conventional waterfall life-cycle model has proven to be ineffective for conventional software development so it is no surprise that developers do not want to use it for expert system development. A more iterative model (e.g., the spiral model) that includes the use of rapid prototyping is being perceived as a better alternative to the waterfall model. "Conventional" software development project often include the use of prototyping, developing better user interfaces, having more user involvement during development, or having developers better understand the problem domain; these are not issues or approaches that are unique to expert system development.

**Small/Simple vs. Large/Complex Systems:** Although some of the systems surveyed are fairly large (e.g., 200 personmonths), they are generally much smaller than dedicated software development projects (e.g., Shuttle MCC, Shuttle flight software, etc.). The systems surveyed seem to be isolated efforts to develop off-line applications for niches for which expert system technology was felt to be very suitable. That is, they were not systems that are not a part of larger software system; though they are often used in conjunction with a large data processing system (e.g., they receive real-time data from a large data processing system). This allowed the expert system developers to work without many of the constraints imposed on larger systems (e.g., tightly controlled configuration management).

**Addressing a Knowledge Engineer Instead of a Programmer:** Although we did not intend to gather information on the experience and background of individual expert system developers, we did learn that several respondents involved in developing expert systems are experts in a problem domain and do not have much programming experience. This fact will be important when considering recommendations (see "Recommendations" on page 23); that is, the recommen-

dations should not assume first-hand knowledge of conventional software V&V techniques.

**Summary:** It may be the case that the above issues are indeed typical of expert system development projects and that they should be addressed when addressing V&V of expert system problems. However, it should be recognized that they are somewhat different than the other issues that are true of all expert systems regardless of their size and who is developing them. This may point to a need to tailor suggestions for V&V of expert systems to considerations such as the size of the expert system, the experience of the developer, whether the system is embedded in a much larger software system, etc.

---

## Recommendations

The recommendations from the survey results are separated into two categories:

### Direct Recommendations

Recommendations in this category are directly supported by the survey results. These recommendations include:

- Develop Requirements for Expert System Verification and Validation
- Address Most Often Encountered Issues
- Recommend a Life Cycle for Expert Systems Development

### Inferred Recommendations

Recommendations in this category can be inferred from the survey results by analyzing relationships among the responses. These recommendations include:

- Address Readability and Modularity Issues
- Address Configuration Management Issue
- Develop Criteria to Classify Expert Systems by Intended Use
- Investigate Applicability of Analysis Tools

Following each general recommendation is an explanation of what was observed in the survey results. After this explanation is a list of specific recommendations which address all the observations. Each specific recommendation in the "Direct Recommendations" section is followed by a list of supporting phrases from "Summary of Results" on page 8.

---

## Direct Recommendations

### Develop Requirements for Expert System Verification and Validation

The major goal of this survey task was to discover and document the current state of the practice in Verification and Validation of Expert Systems. Based on the survey results, it appears that much can be done to improve the practice. The lack of requirements for performing V&V on ESs was manifested in several forms:

- The V&V activities performed were very inconsistent, ranging from none to very many, and the sets of activities performed were very diverse.
- The reliance on expert consultation as the only source of requirements was extremely high.
- The reliance on experts to perform V&V activities on the knowledge base, interface code, and executing systems was very high.
- The low performance levels for many of the expert systems was surprising. Although it is not known what is acceptable reliability for the systems that were surveyed, often the estimated actual reliability was less than the expected reliability. Also, it is unlikely that conventional software systems that exhibited a similar level of performance would gain wide acceptance. (For example, many reported that the ES provides the correct answer less than 90 % of the time. Most conventional software reliability is rated as a series of '9's, e.g., 4 '9's means the correct answer is given > 99.99 % of the time.)

- In those cases where the expected behavior of the system was not strictly defined by expert consultation, a large number of systems relied on prototypes. This is significant because prototype systems receive less V&V than operational systems, but are then used to define the behavior of operational systems.

Each of the above observations can be directly attributed to three factors:

1. There is a general lack of understanding on how to V&V ESs. The wide ranging use of V&V approaches (e.g., each technique being used as the sole technique by at least one project) indicates that there is no clear approach to V&V. That is, it is not known what V&V activities are to be performed, when the activities should be performed, or how the activities can be accomplished. This could, in part, be due to the software experience level of some of the developers.
2. There is little understanding of how requirements for an ES should be generated and documented. It could be argued that this is a development issue, but without documented expected behavior, there is no possibility of performing adequate V&V.
3. A large number of expert systems are prototypes for which V&V receives little consideration.

#### Recommendations

1. Develop recommendations and/or guidelines for Verification and Validation of Expert Systems. (Since such a significant amount of research has been devoted to V&V of traditional software, it may be appropriate to approach this task as a set of modifications to current conventional software V&V requirements.) These guidelines should include the ability for customization based on system size, developer software experience, whether it is stand-alone or a part of a much larger system, etc.

"75% of the respondents indicated that expert consultation was a basis for determining the behavior of the system."

"Most V&V activities relied on comparison with expected results and expert checking"

"In most cases, there was not a separate group to perform V&V"

2. Initial efforts to define V&V requirements should be focused on diagnostic systems, since a large majority of the systems surveyed performed diagnostic services.

"Most ... perform Diagnosis (45%,80) ..."

3. Research the process of converting prototype ESs into operational systems. A large number of respondents indicated that they were either building prototypes for later conversion into operational systems, or building operational systems based on prototypes.

"43% of respondents indicated that prototypes or similar tools were used for the requirements"

"39% of the respondents indicated that the ES was a prototype system."

## Address Most Often Encountered Issues

All of the known issues with performing V&V on Expert Systems were cited at least once in the survey. A small group of issues, however, were cited significantly more often than others and included:

1. Determining test coverage,
2. Knowledge validation,
3. Real-time performance analysis
4. Complexity of the problem

The first two issues are well understood and are active research areas. These research areas should be matured so that they solutions to these issues can be provided.

The issue of real-time performance analysis was briefly discussed earlier (see "Summary of Interview Results" on page 20). Since this issue may most often be interpreted as the inability to get the expert system to run fast enough, and this is not a V&V issue, it is not clear that any recommended action is needed. However, it did appear from the descriptions of the expert systems, that the ability to predict the response time of the system should not be a major issue for current expert systems so it is not felt that any recommendation is needed at this time.

The complexity issue is not as well understood. There is considerable opinion that the types of problems addressed by ESs are significantly harder than the problems addressed by conventional software. Others maintain the apparent difficulty is attributed to the lack of requirements (see above). In either case, there does not seem to be a way to approach the complexity issue without considering it in the context of the readability and modularity issues, as done in "Address Readability and Modularity Issues" on page 26.

### Recommendations

1. Develop tools and/or methods to support the determination of test coverage.

"The known issues most often cited as problems were: test coverage determination (50%,75%) ..."

2. Develop methods and/or tools to support the knowledge validation activity.

"The known issues most often cited as problems were: ... knowledge validation (44%,75%) ..."

3. Develop methods and/or tools to assist in managing problem complexity.

"The known issues most often cited as problems were: ... problem complexity (39%,40%) ..."

## Recommend a Life Cycle for Expert Systems Development

The most common Life Cycle applied to the development of the ESs included in this survey was the Cyclic model. In the Cyclic model, the stages of requirements, design, knowledge base development, and test are repeated until the final system is developed. The testing activities at the end of each cycle (except the last) lead to the refinement of the requirements that will be used in the successive cycle. Several variations, including some with a fixed number of cycles, have been proposed.

A large number of respondents, however, indicated that no attempt was made to follow any model. If no model is being followed, there is little opportunity to apply V&V activities at the appropriate points during development. Clearly, any life cycle

guidelines would be of benefit in these situations. Multiple life-cycle approaches, or a single very flexible life-cycle should be recommended.

#### **Recommendation**

1. Multiple life cycle models, or a single, very flexible life cycle model should be recommended for development of ESs. (The high incidence of prototypes leading to operational systems suggests that the cyclic model should be recommended. Rapid prototyping could be treated as a special case of the cyclic model.)

"The most frequent (40%) Life-Cycle model used is the Cyclic Model ... however, 22% ... stated that no model was followed."

"43% respondents indicated that prototypes or similar tools were used for the requirements"

"(39%,20%) of the respondents indicated that the ES was a prototype system."

---

## **Inferred Recommendations**

### **Address Readability and Modularity Issues**

Readability and modularity were expected to be significant issues, but were not the most frequently cited problems. Further analysis of the survey results indicate that the readability and modularity issues may have been reported as other problems. This analysis includes the following observations:

- As often as not, people chose modularity or readability as problems, but not both. This seems to indicate that many respondents do not see the relationship between the two.
- Similarly, as often as not, people picked test coverage determination without picking modularity, so the apparent relationship between these two issues was not established.
- The lack of reported relationships between the readability, modularity, and test coverage issues is very confusing, implying, for instance, that a rule can be understood but a test scenario for it can not be developed.
- Readability and complexity of the problem were very rarely chosen together. That is, the developer recognizes that the ES was complicated but attributed this complexity either to the problem or to the solution, but not both. It is questionable that the complexity of the problem and the complexity of the solution can be easily distinguished. (The emergence of Object-oriented programming languages is due, in part, to the claim that conventional languages cause programming complexities which are erroneously attributed to problem complexity.)

If the number of times each of these issues were reported are added together, the collection of issues becomes a very frequently cited problem. Since these issues are so closely interrelated, they should be addressed as a single issue. Therefore, the problem of reducing overall complexity (problem/solution) is a very important issue.

#### **Recommendation**

1. Develop methods and/or tools to support the readability, modularity, and problem complexity issue.

## **Address Configuration Management Issue**

Configuration management was an infrequently cited problem. However, the survey results also show that in practice the applied CM, while sometimes quite good, was generally poor (changes to the knowledge base were not well managed). This contradiction is probably due to the high frequency of prototypes and "in development" responses to the survey. While there are certain applications for which CM may never be a significant issue, certainly there are applications for which CM is a very important issue.

### **Recommendation**

1. Identify the differences between CM of conventional software systems and CM of expert systems. It is not immediately obvious that there are differences.

## **Develop Criteria to Classify Expert Systems by Intended Use**

The survey results indicate that there is a very diverse set of applications which are utilizing ES technology. At least the following types of applications exist:

### **Expert Clone**

Provides expert assistance to a human user. The expert is usually available if the ES does not provide the correct results. The major uses of this type of include: education and capture of true institutional knowledge.

### **Expert Assistant**

Allows the user, typically an expert, to concentrate on the more important aspects of the task. These ESs typically serve as filtering mechanisms.

### **Autonomous**

Limited supervision is applied to the ES. In addition to providing filtering, these systems typically develop and execute plans to handle situations.

A subcategory of Autonomous ESs are time critical ESs. These ESs exist primarily because experts can not interpret data efficiently enough to perform the task in the allotted time.

### **Self-modifying autonomous**

Part of the planned execution is to modify its knowledge base to respond to certain situational data. The application of V&V to this type of problem is currently uncertain.

### **Traditional Software Problem**

Some conventional problems (e.g. discrete event simulation). are more conveniently implemented using expert system shells

It is apparent that because of this diversity, a single set of V&V requirements is probably undesirable. Development of classification criteria allows a simplification of ES V&V requirements. In addition to simplification, classification allows the development of requirements to be concentrated on the types of applications of interest.

### **Recommendations**

1. Develop classification criteria to distinguish among expert systems which require different V&V approaches.

2. Concentrate initial V&V requirements definition effort on autonomous systems, since these systems are likely the most critical.

### **Investigate Applicability of Analysis Tools**

A very large number of respondents indicated that experts were the primary source of requirements and verification. Several of the previous recommendations would reduce this dependence, but there is a class of expert system applications for which expert consultation will continue to be the leading source.

#### **Recommendations**

1. Determine if there is a communication problem between the experts and the knowledge engineers / expert system developers.
2. If a communication problem exists, investigate the possibility of representing Knowledge Base in a form that domain experts can easily, yet accurately, understand.



---

## Appendix A. Detailed results

The following table represents the raw data from the survey of expert system developers. Except for questions number 1 and 4<sup>1</sup> there is a column in the table for each question in the survey. The column headers have a number in parentheses corresponding to the question number in the survey. There is also a short mnemonic representing the subject of the question to facilitate cross reference to the correct survey question.

---

### Summary of Developers Responses (part 1)

---

<sup>1</sup> Answers to questions 1 and 4 are not provided because these would identify survey respondent.

#	General Info			Performance Criteria					Requirements					Development Information					
	Field (5)	Type (6)	Role (2)	Perf of Xprt (10)	P/S Performance		P/S Coverage		Formal (13)	Diff. (14)	Expert			Life Cycle (19)	Languages/Tools (20)	Size (22)	Tot. Effort (29)	Detailed Effort (21)	Seas. (24)
					Exp. (11)	Act. (12)	Exp. (8)	Act. (9)			Avail (17)	No. (15)	Agree (16)						
1	a	___E gh j	a	c	b	c	a	d	dE	c	c	b	-	c	CLIPS/C	160 rules	30	60/0/40	b
2	a	___E gh_i	none	f	d	d	g	h	___E_	c	d	c	90	a	KEE/LISP	100 rules, 120 frames, 400 facts, 105 params, 13K simts	50-75	50/0/50	b
3	a	___e gll_	ab	d	c	f	c	d	___b_E_	d	b	e	90	c	CLIPS/C	300 rules, 100 facts	?	90/0/10	b
4	a	A_ d _	a	i	i	i	i	j	___E_	c	-	none	-	a	CLIPS	rules, facts	8	75/0/25	b
5	a	___E g_	a	i	i	i	i	j	___E_	c	c	c	a	b	LISP	matrices, vectors	3	10/80/10	c
6	a	___dfig_	b	c	a	c	b	d	___ll_e	c	e	d	90	b	PROLOG/C	30 rules	18	-	b
7	a	ab_dE_ ll_j	b	b	b	e	b	f	___a_dE_	c	c	c	60	c	LISP/C	rules, frames	24-48	50/30/20	d
8	a	bode_h_	b	c	b	f	b	g	___c e fault matrix	d	c	d	90	c	LISP	rules, frames	200	40/10/50	b
9	a	___def_H_j	a	j	i	i	i	j	___a_dE_	c	d	b	-	b	CLIPS/Hypercard	100 rules, 200 facts	40	40/0/60	b
10	L	___jk	a	b	e	e	c	a	other tools	c	d	c	60	c	C	500 frames, 50 params	96	30/20/50	b
11	b	___K	a	j	c	c	c	d	___E credit scoring	b	c	e	90	a	TIRIS/C	300 rules, 1 frame, 50 params	42	30/0/70	b
12	b	A_	a	b	a	d	a	h	___c E other tools	d	e	e	75	b	LISP	rules, frames, procs	32	50/37/13	c
13	a	___Ef_	a	100%	a	b	b	g	___dE	b	b	d	99	b	Knowledge Tool/ PLI	200 rules	29	40/0/60	b
14	a	___Ef_	coord	d	c	g	25	5	___E	d	d	d	50	a	Knowledge Tool	200 rules	6	80/0/20	b
15	a	A_	a	c	b	i	f	j	___b_dE	b	a	e	95	a	Knowledge Tool	150 rules	1.5	100/0/0	c
16	proj mgmt	___g_k	lead	b	b	b	a	b	___ll_	b	d	b	a	a	Knowledge Director/ Pascal	network, truth tables	4	60/0/40	a
17	a	___E_	a	d	d	d	e	g	___De	c	c	c	b	b	ESE	rules, parameters	25	65/0/35	b
18	a	___E_	a	b	a	a	a	b	___a_e	c	a	a	a	c	Rulemaster/ Ada	rules	-	20/0/80	b
19	a	___E_ ij	a	b	a	d	b	e	___b_de	d	c	d	85	b	CLIPS	> 300 rules	> 56	85/0/15	c
20	a	___E gh j	b	b	h	-	b	a	___E_	c	d	c	70	b	Neupert / C	rules, frames	6	50/0/50	b
21	a	___E gh j	ab	b	i	-	a	a	___b_e	c	d	c	75	b	Neupert / C	rules, frames, facts	30	45/5/50	b
22	comm	A_ d _	ab	b	d	c	d	f	abcdE	c	c	d	90	b	CLIPS	500 rules, 100 facts	12	40/0/60	b
23	a	___l_	a	b	a	b	b	c	___b_de	d	c	d	85	b	CLIPS	< 50 rules	12	85/0/15	c
24	a	___cE_ h_	a	d	a	e	a	f	___De	d	c	e	85	c	LISP	frames	50	30/50/20	c
25	a	___C e gh_	ab	d	b	f	b	g	___dE	c	e	a	b	a	LISP	frames	200	15/70/15	b
26	a	bCaeigh j	ab	b	a	a	c	f	___b_dE	c	d	c	b	b	CLIPS/ C	250 rules, 100 facts	30	45/5/50	b
27	a	___ef_	b	j	i	c	a	h	ab de	c	e	c	90	c	KEE/ LISP	> 10,000 frames	48	40/30/30	b
28	k	___G_	Xprt	c	a	a	a	a	___l_	a	-	a	a	d	ESE/ Pascal	371 rules, 6 frames, 312 params	2	60/0/40	b
29	petro	A_	a	b	d	d	a	b	lit. search	b	-	a	a	a	LISP	rules, frames	3	40/0/60	b

#	General Info			Performance Criteria					Requirements					Development Information					Sens. (24)		
	Field (5)	Type (6)	Role (2)	Perf of Xprt (10)	ES Per- formance		ES Cov- erage		Formal (13)	Diff. (14)	Expert			Dev User (18)	Life Cycle (19)	Languages/ Tools (20)	Size (22)	Tot. Eff. (29)		Detailed Effort (21)	
					Exp. (11)	Act. (12)	Exp. (8)	Act. (9)			Avail (17)	No. (15)	Agree (16)								
30	i	Code hij	a	b	d	e	e	a	... E	-	-	d	b	-	c	c	CLIPS/ Hypercard	350 rules, 200 frames	-	67/0/33	d
31	a	Code	a	c	b	c	c	d	A_e	b	-	d	b	a	b	c	CLIPS	rules, facts, defenplates	1	75/0/25	c
32	i	E	b	f	a	c	b	e	B other docs	d	d	d	c	must	a	d	Gen-X/ C	rules, facts, params	60	85/0/15	c
33	a	E_h	a	c	a	b	a	e	E	e	c	c	c	90	a	c	CLIPS/ C	250 rules, 300 facts, 30 params	2	50/0/50	b
34	k	G	Xprt	c	a	a	a	a	D	a	-	-	d	b	b	d	ESIE/ Pascal	153 rules, 7 frames, 108 params	1	60/0/40	b
35	a	E_ghi	a	e	d	d	g	a	b de	b	d	d	c	90	a	c	ART/ Ada	70 rules, frames, 350 facts, 100 params, stimis	25	60/0/40	b
36	a	E_Hi	a	c	b	c	d	h	-	-	-	-	-	b	c	-	-	-	-	-	-
37	b	C_K	a	100	a	a	e	a	abcde	b	b	b	c	-	b	c	ESIE	216 rules, 45 params	7	70/0/30	b
38	b	C_e_hl	ab	d	c	c	c	a	Ab e	c	-	-	c	85	b	c	ESIE	rules, facts, params	9	90/0/10	b
39	a	d (ll)	ab	c	a	a	c	a	dl	a	a	a	c	b	a	c	LISP	140 frames	-	80/15/5	b
40	a	D	ab	d	e	b	e	a	dl	a	a	a	c	b	a	c	AWK/C	60 rules	-	90/0/10	b
41	a	ll	ab	c	a	a	e	a	de	a	a	a	c	b	a	c	CLIPS/C/TAE	60 rules	-	35/30/35	b
42	a	jk	ab	d	a	a	e	a	de	a	a	a	c	b	a	c	CLIPS/C	100 rules, 1500 frames	-	90/0/10	b
43	i	D	a	f	b	d	b	e	de	d	b	b	c	90	b	c	LISP	50 rules, 12 frames, 3000 stimis	200	50/20/30	e
44	d	J	a	g	f	h	h	i	E	d	d	d	c	90	b	a	other	-	48	67/0/33	a
45	e	A	a	c	b	a	b	c	A	d	-	-	a	90	b	b	KT/PLI	80 factors	5	70/0/30	b
46	d	b_ef	a	c	d	d	e	f	-	b	-	-	a	-	b	-	ESIE/REXX	40 rules, 20 params	21	63/0/37	b
47	f	other	a	c	b	b	a	c	E	b	b	b	d	40	b	d	PROLOG/C	-	9	67/17/17	b
48	e	other	a	e	d	d	e	f	E	b	c	c	c	90	c	b	ESIE	850 rules	21	60/0/40	a
49	d	b_E_i	c	d	c	c	h	i	B	c	b	b	c	60	a	a	ESIE/REXX	80 rules, 50 params	9	50/0/50	b
50	d	A	a	a	a	b	a	c	E	c	c	c	d	80	b	b	KT	-	150	100/0/0	b
51	h	de_i	a	e	b	c	f	h	bode	c	c	c	c	90	a	c	ESIE/REXX	500 frames, 100 methods	15	78/0/22	b
52	d	A	a	a	d	e	i	i	b De	d	d	d	d	50	b	d	PROLOG	70 rules, 5 frames	48	83/0/17	d
53	i	E	a	d	e	i	e	g	E	c	-	-	a	80	b	f	ESIE/LISP	58 params	9	100/0/0	a
54	i	b_ef	a	d	c	c	b	d	abcde	c	-	-	a	80	b	c	ESIE/REXX	-	48	14/14/72	b
55	d	D	a	b	b	i	h	i	E	d	b	b	c	90	b	c	other	400 rules, 30 frames	5	56/0/44	d
56	e	E	a	c	b	b	e	f	b d	c	c	c	c	b	a	b	other	1000 rules, 25 frames, 100 params	21	36/0/64	a
57	e	abcdef hij	a	c	a	a	a	b	E	b	b	b	e	60	a	e	other	-	5	25/25/50	b
58	e	A	a	b	f	f	g	g	A	b	d	d	c	80	c	b	ESIE	500 rules, 400 params	15	50/0/50	b
59	e	I	a	f	f	h	f	g	E	d	d	d	b	-	-	c	C	-	2	62/38/0	b
60	c	A	a	d	e	e	d	g	E	c	b	b	c	70	c	b	ESIE	60 rules, 120 params	2	75/0/25	d

#	General Info			Performance Criteria				Requirements					Development Information							
	Field (5)	Type (6)	Role (2)	Perf of Xpt (10)	ES Performance		ES Coverage		Formal (13)	Diff. (14)	Expert			Dev User (18)	Life Cycle (19)	Languages/Tools (20)	Size (22)	Tot. Eff. (29)	Detailed Effort (21)	Scas. (24)
					Exp. (11)	Act. (12)	Exp. (8)	Act. (9)			Avail (17)	No. (15)	Agree (16)							
61	k	other	a	b	a	a	e	B	E	c	d	c	90	a	b	PROLOG/ REXX	100 rules	48	67/8/25	c
62	d	E	a	b	a	b	d	e	a d	b	c	c	80	a	c	ESB/REXX	50 rules, 50 facts	2	82/0/18	b
63	d	A	a	c	a	b	a	d	B	b		a		b	d	ESB/REXX	360 rules, 260 params	21	78/0/22	b
64	f	G	c	d	c	e	B	i	E	d	b	c	70	a	d	ESB	800 facts, 400 rules, 40 frames	9	100/0/0	b
65	c	ef	a	d	c	i	d	j		a		a	b	a	f	ESB		2	100/0/0	b
RES	65		65	65	65	63	65	65	62	63	53	64	57	64	63	64	57		57	64
TOT	a-30 b-4 c-2 d-8 e-6 f-2	a-13(11) b-10(0) c-9(4) d-16(5) e-26(21) f-11(1)	a-42 b-6 c-17	a-2 b-17 c-16 d-14 e-4 f-4	a-21 b-14 c-8 d-9 e-4 f-3	a-11 b-10 c-11 d-9 e-8 f-4	a-15 b-12 c-6 d-5 e-12 f-3	a-3 b-3 c-8 d-3 e-14 f-17	a-12(4) b-19(4) c-6(0) d-27(4) e-49(30) f-6(0)	a-7 b-15 c-25 d-15 e-1	a-6 b-10 c-15 d-17 e-5	a-10 b-6 c-30 d-12 e-6	a-6 b-11 c-39 d-75% e-30% f-99%	a-23 b-30 c-11	a-5 b-12 c-25 d-14 e-2 f-2	CLIPS-13 ESB-17 KT-5 LISP-10 C-14 REXX-7 PROLOG-3 LISP-2	rules-47 235.1(35) frames-24 872.4(15) facts-14 248.0(10) params-18 120.5(15) stimis-3 800(2) methods-1 100(1)	33.1 range 1-200	KB = 60.7% IE = 7.7% IC = 31.6%	a-5 b-44 c-9 d-5 e-1

ORIGINAL PAGE IS  
OF POOR QUALITY

---

## Summary of Developers Responses (part 2)

#	V&V Activities Performed										V&V Issues Encountered								
	During Devel. (28)	% V&V (30)	(detail) V&V of				Expert DMF (23)	IV&V (31)	After Devel. (33)	Effort (32)	Oper/Proble (3)	Critical (37)	Encounter (36)	Cer-tainty (7)	CM (34)	V&V DMF (35)	System Use		Reli-able (38)
			KS (25)	IE (26)	IC (27)	Expected (39)											Actual (40)		
1	bc_	15	-	-	b_d	c		a	bode	3%	ab	de	d g i	b	d	c	8	0/50/30/20	h
2	bc_	25	b_d	-	b_e	c		b	bod	-	a	c e	ab _ g _	b	b	c	100	5/75/10/10	b
3	b_	?	b_e	-	c_	d		b	f	-	a	c_	d _ gh _	c	c	d	7	0/43/14/43	h
4	c_	20	b_	-	b_	c		b	-	-	b	e	d _ h _	b	-	c	-	-	b
5	-	-	-	-	-	c		b	-	-	b	a d	gh _	b	-	-	-	-	i
6	c_	35	d_	-	a_	c		b	d	-	b	b	a _ f _	b	a	b	6	15/55/15/15	e
7	bc_e	20	def	-	e_	d		b	bode	-	a	abc	b_d _ hi _	b	-	d	-	-	i
8	c_	2	a_	-	a_	d		b	c_	-	ab	a_c_	bc _ h _	b	a	f	-	0/80/20/0	b
9	bod	20	b_de	-	bod	c		-	bod	-	b	c_	bod _ g _ i _	b	-	-	5	-	b
10	c_	5	c_	-	d_	b		b	c_	f	a	d	h _ k _	b	user updt	f	50	0/10/0/90	d
11	b_	20	od	-	a_	b		b	e	-	b	d	b_d _	b	f	d	1000	-	b
12	bode	60	e_	-	d_	d		a	-	5%	b	e	-	b	e	c	2000	0/10/90/0	d
13	bc_	15	b_d	-	b_d	c		a	bod	THD	b	c_e	_____k	b	d	e	150-300	?	h
14	bod	20	bodef	-	bode	d		b	bode	-	b	a	abcde _ hi _ k	a	-	c	3	0/0/10/90	d
15	bod	50	b_e	-	a_	d		b	-	-	b	abc	bod _ hi _	c	-	c	20-40	-	h
16	d_	25	e_	-	b_d	b		b	bc_	-	a	cde	-	b	d	e	-	-	h
17	bod	10	b_de	-	b_d	d		b	bc	-	b	d	d f h _	b	-	e	10	-	i
18	bc_	20	od	-	c_	c		a	bod	-	a	c_	_____g _ k	b	-	c	-	-	a
19	bc	30	d	-	b_	c		b	bod	-	a	ab e	b d ghij	b	b	c	-	-	h
20	e_	40	b_def	-	b_de	d		b	bodef	-	a	cde	bod _ h _ j	b	b	de	10-15	?	i
21	ef	40	b_def	-	b_de	d		b	bodef	-	a	de	b d h j	b	b	de	10	-	h
22	bcd	5	b_d	-	a_	c		b	bod	-	a	c	abcd f	b	ab	d	500	0/40/50/10	h
23	bc_	80	d_	-	b_	c		b	bod	-	a	cde	b d ghij	b	b	d	-	-	h
24	c_	10	b_f	-	c_	c		b	c	-	b	c	abcd fgh	b	b	d	-	-	b
25	c_	10	f	-	e_	c		b	a	-	c	-	bod fghij	b	-	f	-	-	b
26	bod	10	b_d	-	bod	c		b	bod	-	a	b	abcd g	b	b	nune	1	-	d
27	c	5	d	-	d	d		b	-	-	b	c	b d g j	b	-	d	> 20	-	h
28	b_e	5	b_d f	-	b_e	c		a	c	-	a	c_	_____k	b	d	b	40	60/30/10/0	b
29	bc_	10		-	a_	b		b	bc	-	a	c_	a _ g _	a	d	b	10	0/20/0/80	d

ORIGINAL PAGE IS  
OF POOR QUALITY

#	V&V Activities Performed										V&V Issues Encountered								
	During Devel. (28)	% V&V (30)	(detail) V&V of				Expert Diff. (23)	IV&V (31)	After Devel. (33)	Effort (32)	Oper/ Proto (3)	Criti- cal (37)	Encounter (36)	Cer- tainty (7)	CM (34)	V&V Diff. (35)	System Use		Reli- able (38)
			KS (25)	IE (26)	IC (27)	Expected (39)											Actual (40)		
30	cd	30	de	a	d	c	b	bode	-	b	bde	b	b	e	c	4	-	h	
31	cd	30	de	a	d	c	b	-	-	b	b	b	a	c	-	c	-	h	
32	cd	-	de	a	d	c	-	cd f	-	b	b	b	b	d	d	> 2000	-	b	
33	bcd	15	de	a	d	c	b	bde	-	b	b	b	cd	b	b	c	-	d	
34	bde	5	de	a	d	c	a	cd	-	a	c	b	k	d	d	b	14	60/30/10/0	b
35	cd	75	de	a	d	c	b	ef	-	a	c	b	bc	f	b	c	100	-	-
36	-	-	-	-	-	-	-	bde	-	a	b	b	bde	gh	-	d	5-10	30/60/10/0	b
37	bcd	30	de	a	d	c	a	bde	7%	a	c	a	b	-	-	c	20-40	-	a
38	b	20	de	a	d	c	a	bde	-	a	e	b	de	a	-	d	100	-	h
39	bcd	25	de	a	d	b	b	bcd	-	b	c	b	b	bd	b	b	-	-	d
40	bcd	25	de	a	d	b	b	bcd	-	a	c	b	b	bd	b	b	18	0/100/0/0	d
41	bcd	25	de	a	d	b	b	bcd	-	b	c	b	b	bd	b	b	-	-	d
42	bcd	25	de	a	d	b	b	bcd	-	b	c	b	b	bd	b	b	-	-	d
43	a	10	de	a	d	d	b	-	-	b	b	b	abc	h	-	c	100	-	a
44	d	2	de	a	d	e	b	-	-	b	e	a	b	b	b	c	25-50	0	d
45	e	21	de	a	d	d	b	-	-	a	d	i	b	d	c	c	5-10	9	a
46	d	5	de	a	d	b	b	-	-	a	b	d	b	d	b	b	> 500	7	b
47	b	30	de	a	d	d	b	-	-	a	e	e	a	d	c	c	> 500	9	a
48	e	5	de	a	d	b	b	-	-	a	d	b	b	d	c	c	10-25	7	c
49	e	9	de	a	d	d	b	-	-	a	c	h	b	d	d	d	25-50	8	d
50	bde	9	de	a	d	d	a	bcd	30	a	b	a	c	h	d	d	> 500	2	d
51	e	5	de	a	d	c	b	-	-	a	b	b	-	b	d	b	> 500	8	d
52	bde	30	de	a	d	d	a	bcd	5	a	d	i	b	d	c	c	150-250	8	d
53	d	9	de	a	d	c	b	-	-	a	c	a	b	a	b	b	> 500	8	b
54	bde	5	de	a	d	c	a	bde	5	a	e	cd	b	d	d	d	> 500	8	a
55	bde	21	de	a	d	d	a	-	1	b	c	d	gh	b	e	e	< 5	0	c
56	d	5	de	a	d	c	b	-	-	a	c	b	b	d	b	b	10-25	8	a
57	bde	21	de	a	d	b	b	-	-	a	a	-	-	c	a	a	> 500	9	a
58	bde	5	de	a	d	b	a	c	1	a	d	d	-	b	b	b	10-25	1	d
59	a	2	de	a	d	c	b	-	-	a	e	c	h	a	b	b	5-10	0	e
60	b	5	de	a	d	c	a	c	1	b	d	e	a	-	c	c	< 5	8	c

#	V&V Activities Performed										V&V Issues Encountered								
	During Devel. (28)	% V&V (30)	(detail) V&V of				Expert Intf. (23)	IV&V (31)	After Devel. (33)	Effort (32)	Oper/Proto (3)	Critical (37)	Encounter (36)	Cer-tainty (7)	CM (34)	V&V Intf. (35)	System Use		Reli-able (38)
			KS (25)	IK (26)	IC (27)	Expected (39)											Actual (40)		
61	a, d	9	a	b	c	b				a	c	bc	b	d	d	10-25	2	d	
62	b, e	5	d	e	b	b				a	b	d	b	e	b	100-150	9	a	
63	b, de	2	de	e	b	b				b	d	b	b	b	a	25-50	6	b	
64	b	5	b		b	a		b	1	a	c	h	b	d	b	> 500	7	b	
65	b, e	5	d		a	b				a	e	b	a	b	a	10-25	1	d	
RESP	63	62	63	35	58	64	62	43	10	65	64	62	64	45	63	50	36	64	
TOT	a-2 b-38 c-32 d-28 e-18	23.97% (range 2%-80%) 1-3 2-10 3-4 5-3 6-2	a-3 b-28 c-15 d-42 e-39 f-9	a-17 b-2 c-3 d-10 e-5	a-7 b-25 c-12 d-29 e-18	a-1 b-15 c-30 d-17 e-1	a-15 b-47	a-1 b-29 c-36 d-25 e-13 f-5	5% 1-4 3-2 6-1	a-38 b-24 c-3	a-7 b-14 c-29 d-16 e-18	a-13 b-31 c-18 d-27 e-5 f-7 g-25 h-24 i-14 j-8 k-6	a-7 b-34 c-3	a-5 b-18 c-1 d-22 e-3 f-1	a-3 b-16 c-20 d-16 e-3 f-4	219 range 1-2000	12.14/43.07/ 19.21/25.57 0-3 1-2 2-2 6-1 7-3 8-7 9-4	a-9 b-15 c-3 d-18 e-2 h-12 i-4	

ORIGINAL PAGE IS  
OF POOR QUALITY



## Summary of Users Responses

#	General Info			Performance Criteria					Requirements		
	Field (5)	Type (6)	Role (2)	Performance			Coverage		Expert Agree (21)	Exp User (12)	Dev User (13)
				Xprt (9)	Exp. (10)	Act. (11)	Exp. (7)	Act. (8)			
1	a	_b_E_Hi_	c	c	b	c	d	h	c	a	c
2	a	___e_Hij_	a	c	a	b	f	a	90%	a	b
3	a	_C_e_gh_	a	d	d	f	e	f	90%	a	c
4	a	___E_____	a	c	b	< 40	i	< 40	-	a	a
5	a	_c_e_gh_j_	a	d	c	d	d	f	90%	a	a
RESP	5	5	5	5	5	5	5	5	4	5	5
TOT	a-5	b-1(0) c-2(1) e-5(2) g-2(0) h-4(2) i-2(0) j-2(1)	a-4 c-1	c-3 d-2	a-1 b-2 c-1 d-1	b-1 c-1 d-1 f-1 h-1	d-2 e-1 f-1 i-1	a-1 f-2 h-2	c-1 90%-3	a-5	a-2 b-1 c-2

#	V&V Activities Performed				V&V Issues Encountered					
	After Devel. (20)	Effort (19)	Oper/ Proto (3)	Crit- ical (15)	Encounter (23)	V&V Diff. (22)	System Use		Reli- able (16)	User Trust (14)
							Expected (17)	Actual (18)		
1	_bc_e	-	a	b	_b_defgh_	d	5-10	20 70 10 0	b	a_de
2	_b_de	3	a	d	_c_h_j_	d	5	-	h	_b_d_
3	_bcde	20	a	e	_b_d_____	d	10-15	0 80 20 0	f	_b_d_
4	-	-	b	c	-	-	100	-	h	__d_
5	_bcd_	24	a	c	_b_d_____	d	10-15	-	d	_b_d_
RESP	4	3	5	5	4	4	5	2	5	5
TOT	b-4 c-3 d-4 e-3	15.67 range 3-24	a-4 b-1	b-1 c-2 d-1 e-1	b-3 c-1 d-3 e-1 f-1 g-1 h-2 j-1	d-4	27.5 range 5-100	10 75 15 0	b-1 d-1 f-1 h-2	a-1 b-3 d-5 e-1

ORIGINAL PAGE IS  
OF POOR QUALITY

## Appendix B. Expert Systems Evaluation Questionnaire (Developer)

By filling out this NASA funded questionnaire, you can help define the state-of-the-practice in the formal evaluation of Expert Systems on current NASA and industry applications. The information that you provide will be merged with the information from all other surveyed projects for the purpose of recommending future research and development activities. Individual responses are used solely as input to this information merging process. Each survey participant will be sent a copy of the final survey results.

Expert System applications are becoming more prevalent in fields where proper functioning is essential, such as the aerospace, medical, and financial industries. It is widely claimed that Expert Systems are not as rigorously evaluated as traditional software because of unique, unresolved evaluation issues. To ensure the continued and safe deployment of Expert Systems into critical areas, adequate evaluation techniques which address these issues must be developed and performed.

### Instructions

The following questions concern your experiences with an Expert System, either as a developer or as the manager of the development effort. Feel free to indicate your answers in any way you like. Some of the choices on the multiple choice questions have places to fill in additional information; please indicate the choice and include the additional information, if possible. If you have any comments about the questions or your answers, please write them in the left margin.

Analysis of the responses may indicate that further discussion is required for complete understanding of the issues encountered during the evaluation process. Discussions will be held either as short one-on-one meetings or by telephone. Would you be available, at your convenience, to discuss the evaluation process in more detail?

**Yes** I am available for discussions.

Name \_\_\_\_\_

Phone \_\_\_\_\_

**No** I am not available for discussions.

If you have any questions regarding this questionnaire, please contact Keith Kelley at (713) 282-7303. If possible, please return completed questionnaires within one week of receipt to:

Keith Kelley  
MC 6606  
IBM Federal Sector Division  
3700 Bay Area Blvd.  
Houston, Tx. 77058-1199

## Definitions

### Certainty factors

Some problems require the use of certainty factors (also called probabilities, or fuzzy logic) in their processing. Facts which contain certainty factors have the form: "if a is true, then there is an x% chance that b is true."

### Expert

The person who provides the knowledge that is to be captured in the Expert System.

### Inference engine

Processes the knowledge structures to infer a set of output facts from a set of input facts. Examples of commercial systems are CLIPS and ESE.

### Interface code

Used to supplement the inference process. Examples are interfacing the inference engine to a device, and performing arithmetic calculations.

### Knowledge structures

Declarative part of the Expert System which represents the knowledge (typically called the Knowledge Base). Examples are frames and rules.

### Problem space

The total number of cases which could potentially be addressed by the Expert System.

### Problem space coverage

The percentage of the problem space that is addressed by the Expert System. For example, if the Expert System is supposed to be able to diagnose 100 malfunctions, but the total number of malfunctions is known to be 200, the problem space coverage is 50%.

## Questions

1. What is the name of the Expert System you were/are involved with?  
\_\_\_\_\_
2. Were you a developer of the Expert System or the manager of the development organization?
  - a. Developer of Expert System
  - b. Manager of Expert System development organization
  - c. Other \_\_\_\_\_
3. Is the Expert System operational or is it a prototype?
  - a. Operational system
  - b. Prototype system
4. Briefly describe what the expert system does.  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

5. What field does the problem belong to?

- |                        |                |
|------------------------|----------------|
| a. Aerospace           | g. Medical     |
| b. Financial           | h. Personnel   |
| c. Information Systems | i. Research    |
| d. Hardware            | j. Service     |
| e. Manufacturing       | k. Software    |
| f. Marketing           | l. Other _____ |

6. Which of the following items best describes the kind of problem the Expert System addresses? Please indicate primary purpose with a "\*" and check all other applicable purposes (if any).

- a. Design - Configuring objects under constraints
- b. Repair - Executing plans to administer prescribed remedies
- c. Control - Governing overall system behavior
- d. Planning - Designing actions
- e. Diagnosis - Inferring system malfunctions from observables
- f. Debugging - Prescribing remedies for malfunctions
- g. Prediction - Inferring likely consequences of given situations
- h. Monitoring - Comparing observations to expected outcomes
- i. Instruction - Diagnosing, debugging, and repairing behavior
- j. Interpretation - Inferring situation descriptions from sensor
- k. Classification - Categorizing objects by properties data

7. Does the Expert System include certainty factors?

- a. Yes
- b. No
- c. I don't know

8. How much of the problem space is the Expert System expected to cover?

- |               |                 |
|---------------|-----------------|
| a. 100%       | f. 60% to 80%   |
| b. > 99%      | g. 40% to 60%   |
| c. 95% to 99% | h. Other _____% |
| d. 90% to 95% | i. I don't know |
| e. 80% to 90% |                 |

9. What is your *estimate* of the problem space coverage actually provided by the Expert System?

- |                     |                 |
|---------------------|-----------------|
| a. Same as expected | f. 80% to 90%   |
| b. 100%             | g. 60% to 80%   |
| c. > 99%            | h. 40% to 60%   |
| d. 95% to 99%       | i. Other _____% |
| e. 90% to 95%       | j. I don't know |

Questions 10 through 12 are concerned with the percentage of problems within the problem space (covered by the Expert System) that are answered correctly.

10. If human experts currently perform (or previously performed) the task, how often is the expert(s) expected to give the correct answer?

- |                                |                 |
|--------------------------------|-----------------|
| a. Task not performed by human | f. 80% to 90%   |
| b. "Correct" defined by expert | g. 60% to 80%   |
| c. > 99%                       | h. 40% to 60%   |
| d. 95% to 99%                  | i. Other _____% |
| e. 90% to 95%                  | j. I don't know |

11. How often is the Expert System expected to provide the correct answer?

- |               |                 |
|---------------|-----------------|
| a. 100%       | f. 60% to 80%   |
| b. > 99%      | g. 40% to 60%   |
| c. 95% to 99% | h. Other _____% |
| d. 90% to 95% | i. I don't know |
| e. 80% to 90% |                 |

12. What is *your estimate* of how often the Expert System *actually* provides the correct answer?

- |               |                 |
|---------------|-----------------|
| a. 100%       | f. 60% to 80%   |
| b. > 99%      | g. 40% to 60%   |
| c. 95% to 99% | h. Other _____% |
| d. 90% to 95% | i. I don't know |
| e. 80% to 90% |                 |

13. What was the basis for determining how the system was to behave? Please indicate the primary basis with a '\*' and check all other applicable basis (if any).

- A pre-existing document \_\_\_\_\_
- A requirements document completed as part of development.
- Some other developed document \_\_\_\_\_
- A prototype of the system
- Expert consultation
- Other \_\_\_\_\_

14. How difficult was it to develop the original concept of what the system was supposed to do?

- |            |               |
|------------|---------------|
| a. Trivial | d. Hard       |
| b. Easy    | e. Impossible |
| c. Medium  |               |

15. Was more than one expert consulted during the development of the system?

- |                                   |                               |
|-----------------------------------|-------------------------------|
| a. System was developed by expert | c. Multiple experts with lead |
| b. Single expert                  | d. Committee of experts       |
|                                   | e. Other _____                |

16. If more than one expert was available for consulting, how often did the experts agree on what results the Expert System was supposed to provide?
- a. A single expert was involved
  - b. Always agree
  - c. Agree \_\_\_\_\_% of the time.
  - d. Regular
  - e. Occasional
  - f. None
17. If the system was not developed by the expert, how much interaction was there between the expert(s) and the development team?
- a. System was developed by expert
  - b. Constant
  - c. Frequent
  - d. Regular
  - e. Occasional
  - f. None
18. Was the developer(s) part of the user organization?
- a. Yes
  - b. No
  - c. Some developers were in the user organization
19. Please indicate which development model was used for developing the Expert System.
- a. Requirements gathering preceded Design, Implementation, and Test (Traditional waterfall life-cycle).
  - b. Requirements gathered before development of a prototype. A second requirements activity preceded Design, Implementation, and Test.
  - c. Repetition of the Requirements, Design, Rule Generation, and Prototyping phases until production system (final prototype) was developed.
  - d. No effort was made to follow a particular model.
  - e. Other \_\_\_\_\_
20. What was the primary language/tool for each part of the Expert System?
- a. Knowledge Structures \_\_\_\_\_
  - b. Inference Engine \_\_\_\_\_
  - c. Interface Code \_\_\_\_\_
21. What percentage of the total development effort was dedicated to each part of the Expert System?
- a. Knowledge Structures \_\_\_\_\_%
  - b. Inference Engine \_\_\_\_\_% (If an Expert System Shell was used, this value should be 0%.)
  - c. Interface Code \_\_\_\_\_%

22. Since Knowledge Bases can be written using several type of Knowledge Structures, please indicate how many of the following structures were used. If another type of structure was used, please describe it and how many were used.
- |                 |                             |
|-----------------|-----------------------------|
| a. Rules _____  | d. Parameters _____         |
| b. Frames _____ | e. Statements _____         |
| c. Facts _____  | f. Other (#) _____ of _____ |
23. Aside from any difficulties in developing the original concept, how difficult was it to express the behavior (through the Knowledge Structures) of the expert?
- |            |               |
|------------|---------------|
| a. Trivial | d. Hard       |
| b. Easy    | e. Impossible |
| c. Medium  |               |
24. When changes were made to the knowledge structures, how often did some unexpected result occur?
- |                 |            |
|-----------------|------------|
| a. Never        | d. Usually |
| b. Occasionally | e. Always  |
| c. Frequently   |            |

Questions 25 through 28 are concerned with the evaluation activities performed during development.

25. What evaluation activities were performed on the knowledge Structures? (indicate any that apply)
- |                                |  |
|--------------------------------|--|
| a. No evaluation was performed | d. Checked by expert(s)                      |
| b. Desk checking               | e. Structural testing (e.g. cover all rules) |
| c. Formal inspections          | f. Other _____                               |
26. What evaluation activities were performed on the Inference Engine? (indicate any that apply)
- |                                |                       |
|--------------------------------|-----------------------|
| a. No evaluation was performed | d. Structural testing |
| b. Desk checking               | e. Other _____        |
| c. Formal inspections          |                       |
27. What evaluation activities were performed on the Interface Code? (indicate any that apply)
- |                                |  |
|--------------------------------|--|
| a. No evaluation was performed | d. Structural testing (branch or path) |
| b. Desk checking               | e. Other _____                         |
| c. Formal inspections          |  |

28. What testing activities were performed on the executing system? (indicate any that apply)
  - a. No evaluation was performed
  - b. Checked by expert(s)
  - c. Compared with expected results
  - d. Structural testing (e.g. cover all rules)
  - e. Other \_\_\_\_\_
29. How much effort was expended in developing the system, including evaluation activities performed by the developers? \_\_\_\_\_ person/months.
30. How much of the development effort was spent on evaluation? \_\_\_\_\_ %.
31. Did a separate organization evaluate the Expert System before it was delivered to the users?
  - a. Yes, there was a separate evaluation organization.
  - b. No, there was not a separate evaluation organization.
32. If there was a separate evaluation team, how much effort was expended by the team in evaluating the correctness of the Expert System? \_\_\_\_\_ person/months.
33. What testing activities were performed on the executing system before the system was delivered to the users? (indicate any that apply)
  - a. No evaluation was performed
  - b. Checked by expert(s)
  - c. Compared with expected results
  - d. User acceptance
  - e. System run in parallel
  - f. Other \_\_\_\_\_
34. How were changes to the Expert System distributed to the users?
  - a. User updated system at developer's direction
  - b. Developers made changes to users' system
  - c. Untested system distributed to users
  - d. Tested system distributed to the users
  - e. Configuration management group distributes system
  - f. Other \_\_\_\_\_
35. Compared to conventional software testing efforts, how difficult was the evaluation of the Expert System?
  - a. Trivial
  - b. Easy
  - c. Medium
  - d. Hard
  - e. Impossible
  - f. No evaluation was done



36. Many people feel that some development issues are more of a problem with Expert Systems than with conventional systems. Which (if any) of the following were problems during implementation or test of this Expert System?
- a. Understandability and readability of knowledge structures
  - b. Determining test coverage for knowledge structures
  - c. Modularity/Design of knowledge structures
  - d. Knowledge validation
  - e. Analysis of Certainty Factors
  - f. Validating the inference engine
  - g. Real-time performance analysis
  - h. Complexity of the Problem
  - i. Certification
  - j. Configuration Management
  - k. Other \_\_\_\_\_
37. How reliable is the Expert System required to be?
- a. Trusted with human life
  - b. Trusted with mission objectives
  - c. As reliable as the expert
  - d. Assists the expert
  - e. Assists the user
  - f. Other \_\_\_\_\_
38. Does the Expert System seem to be more reliable or less reliable than conventional systems that are in use?
- a. Significantly more reliable
  - b. More reliable
  - c. Slightly more reliable
  - d. Similar reliability
  - e. Slightly less reliable
  - f. Less reliable
  - g. Significantly less reliable
  - h. No comparison is available
  - i. I don't know
39. How many people are expected to make use of the Expert System?  
\_\_\_\_\_
40. How frequently are the (expected) users actually using the system? (Numbers may add up to more than 100% if the actual number of users is greater than the expected users.)
- a. \_\_\_\_\_% use the system more than expected
  - b. \_\_\_\_\_% use the system about as much as expected
  - c. \_\_\_\_\_% use the system less than expected
  - d. \_\_\_\_\_% do not use the system

## Appendix C. Expert Systems Evaluation Questionnaire (User)

By filling out this NASA funded questionnaire, you can help define the state-of-the-practice in the formal evaluation of Expert Systems on current NASA and industry applications. The information that you provide will be merged with the information from all other surveyed projects for the purpose of recommending future research and development activities. Individual responses are used solely as input to this information merging process. Each survey participant will be sent a copy of the final survey results.

Expert System applications are becoming more prevalent in fields where proper functioning is essential, such as the aerospace, medical, and financial industries. It is widely claimed that Expert Systems are not as rigorously evaluated as traditional software because of unique, unresolved evaluation issues. To ensure the continued and safe deployment of Expert Systems into critical areas, adequate evaluation techniques which address these issues must be developed and performed.

### Instructions

The following questions concern your experiences with an Expert System, either as a user or as the manager of a department that uses Expert System. Feel free to indicate your answers in any way you like. Some of the choices on the multiple choice questions have places to fill in additional information; please indicate the choice and include the additional information, if possible. If you have any comments about the questions or your answers, please write them in the left margin.

Analysis of the responses may indicate that further discussion is required for complete understanding of the issues encountered during the evaluation process. Discussions will be held either as short one-on-one meetings or by telephone. Would you be available, at your convenience, to discuss the evaluation process in more detail?

**Yes** I am available for discussions.

Name \_\_\_\_\_

Phone \_\_\_\_\_

**No** I am not available for discussions.

If you have any questions regarding this questionnaire, please contact Keith Kelley at (713) 282-7303. If possible, please return completed questionnaires within one week of receipt to:

Keith Kelley  
MC 6606  
IBM Federal Sector Division  
3700 Bay Area Blvd.  
Houston, Tx. 77058-1199

## Definitions

### Expert

The person who provides the knowledge that is to be captured in the Expert System.

### Inference engine

Processes the knowledge structures to infer a set of output facts from a set of input facts. Examples of commercial systems are CLIPS and ESE.

### Knowledge structures

Declarative part of the Expert System which represents the knowledge (typically called the Knowledge Base). Examples are frames and rules.

### Problem space

The total number of cases which could potentially be addressed by the Expert System.

### Problem space coverage

The percentage of the problem space that is addressed by the Expert System. For example, if the Expert System is supposed to be able to diagnose 100 malfunctions, but the total number of malfunctions is known to be 200, the problem space coverage is 50%.

## Questions

1. What is the name of the Expert System you were/are involved with?  
\_\_\_\_\_
2. Are you a user of the Expert System or the manager of a department which uses the Expert System?
  - a. User of the Expert System
  - b. Manager of a department using the Expert System
  - c. Other \_\_\_\_\_
3. Is the Expert System operational or is it a prototype?
  - a. Operational system
  - b. Prototype system
4. Briefly describe what the expert system does.  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
5. What field does the problem belong to?

a. Aerospace	g. Medical
b. Financial	h. Personnel
c. Information Systems	i. Research
d. Hardware	j. Service
e. Manufacturing	k. Software
f. Marketing	l. Other _____

6. Which of the following items best describes the kind of problem the Expert System addresses? Please indicate primary purpose with a '\*' and check all other applicable purposes (if any).

- a. Design - Configuring objects under constraints
- b. Repair - Executing plans to administer prescribed remedies
- c. Control - Governing overall system behavior
- d. Planning - Designing actions
- e. Diagnosis - Inferring system malfunctions from observables
- f. Debugging - Prescribing remedies for malfunctions
- g. Prediction - Inferring likely consequences of given situations
- h. Monitoring - Comparing observations to expected outcomes
- i. Instruction - Diagnosing, debugging, and repairing behavior
- j. Interpretation - Inferring situation descriptions from sensor data
- k. Classification - Categorizing objects by properties

7. How much of the problem space is the Expert System expected to cover?

- a. 100%
- b. > 99%
- c. 95% to 99%
- d. 90% to 95%
- e. 80% to 90%
- f. 60% to 80%
- g. 40% to 60%
- h. Other \_\_\_\_\_%
- i. I don't know

8. What is *your estimate* of the problem space coverage actually provided by the Expert System?

- a. Same as expected
- b. 100%
- c. > 99%
- d. 95% to 99%
- e. 90% to 95%
- f. 80% to 90%
- g. 60% to 80%
- h. 40% to 60%
- i. Other \_\_\_\_\_%
- j. I don't know

Questions 9 through 11 are concerned with the percentage of problems within the problem space (covered by the Expert System) that are answered correctly.

9. If human experts currently perform (or previously performed) the task, how often is the expert(s) expected to give the correct answer?

- a. Task not performed by human
- b. "Correct" defined by expert
- c. > 99%
- d. 95% to 99%
- e. 90% to 95%
- f. 80% to 90%
- g. 60% to 80%
- h. 40% to 60%
- i. Other \_\_\_\_\_%
- j. I don't know

10. How often is the Expert System expected to provide the correct answer?

- a. 100%
- b. > 99%
- c. 95% to 99%
- d. 90% to 95%
- e. 80% to 90%
- f. 60% to 80%
- g. 40% to 60%
- h. Other \_\_\_\_\_%
- i. I don't know

11. What is *your estimate* of how often the Expert System actually provides the correct answer?
 

a. 100%	f. 60% to 80%
b. > 99%	g. 40% to 60%
c. 95% to 99%	h. Other _____ %
d. 90% to 95%	i. I don't know
e. 80% to 90%	
  
12. Was the expert(s) a member of the user organization?
 

a. Yes	c. User organization provided some expertise
b. No	
  
13. Was the developer(s) of the Expert System part of the user organization?
 

a. Yes	c. Some development provided by user organization
b. No	
  
14. Why do you believe the results that the system gives?
 

a. Expert says it is correct	e. User acceptance
b. Participated in evaluation	f. I don't trust the results
c. Someone I trust did evaluation	g. Other _____
d. Personal use and checking	
  
15. How reliable is the Expert System required to be?
 

a. Trusted with human life	d. Assists the expert
b. Trusted with mission objectives	e. Assists the user
c. As reliable as the expert	f. Other _____
  
16. Does the Expert System seem to be more reliable or less reliable than conventional systems that are in use?
 

a. Significantly more reliable	f. Less reliable
b. More reliable	g. Significantly less reliable
c. Slightly more reliable	h. No comparison is available
d. Similar reliability	i. I don't know
e. Slightly less reliable	
  
17. How many people are expected to make use of the Expert System?  
 \_\_\_\_\_

18. How frequently are the (expected) users actually using the system? (Numbers may add up to more than 100% if the actual number of users is greater than the expected users.)
- a. \_\_\_\_\_ % use the system more than expected
  - b. \_\_\_\_\_ % use the system about as much as expected
  - c. \_\_\_\_\_ % use the system less than expected
  - d. \_\_\_\_\_ % do not use the system

If you were not involved with evaluating the Expert System, please leave the remaining questions unanswered.

19. How much effort was expended by the evaluation team in evaluating the correctness of the Expert System? \_\_\_\_\_ person/months.
20. What testing activities were performed on the executing system before the system was delivered to the users? (indicate any that apply)
- a. No evaluation was performed
  - b. Checked by expert(s)
  - c. Compared with expected results
  - d. User acceptance
  - e. System run in parallel
  - f. Other \_\_\_\_\_
21. If more than one expert was available for consulting, how often did the experts agree on what results the Expert System is supposed to provide?
- a. No expert was involved
  - b. A single expert was involved
  - c. Always agree
  - d. Agree \_\_\_\_\_ % of the time.
22. Compared to conventional software testing efforts, how difficult was the evaluation of the Expert System?
- a. Trivial
  - b. Easy
  - c. Medium
  - d. Hard
  - e. Impossible
23. Many people feel that some development issues are more of a problem with Expert Systems than with conventional systems. Which (if any) of the following were problems during testing of the Expert System?
- a. Understandability and readability of knowledge structures
  - b. Determining test coverage for knowledge structures
  - c. Modularity/Design of knowledge structures
  - d. Knowledge validation
  - e. Analysis of Certainty Factors
  - f. Validating the inference engines
  - g. Real-time performance analysis
  - h. Complexity of the Problem
  - i. Certification
  - j. Other \_\_\_\_\_